

# Type I Restriction Systems: Sophisticated Molecular Machines (a Legacy of Bertani and Weigle)

NOREEN E. MURRAY\*

*Institute of Cell and Molecular Biology, University of Edinburgh,  
 Edinburgh EH9 3JR, United Kingdom*

<b>GENERAL INTRODUCTION.....</b>	<b>413</b>
Host-Controlled Modification—A Historical Perspective .....	413
Nomenclature and Classification of R-M Systems.....	413
Nomenclature .....	413
Classification .....	414
<b>INTRODUCTION TO TYPE I R-M SYSTEMS.....</b>	<b>414</b>
General Characteristics .....	414
The Family Concept.....	414
<b>ENZYMES .....</b>	<b>415</b>
Introduction.....	415
Specificity Subunit—HsdS.....	415
Variable sequences.....	415
Conserved sequences .....	415
Imposition of symmetry .....	415
Association of subunits .....	417
DNA specificity.....	417
Molecular model .....	417
HsdM needed for DNA binding.....	418
Modification Subunit—HsdM.....	418
Active site.....	418
Modification Enzyme.....	419
Discrimination of methylation state.....	419
DNA binding.....	419
Base flipping.....	419
Restriction Subunit—HsdR.....	420
R-M Complex .....	420
DNA binding.....	420
Recognition of methylation state .....	420
Communication .....	421
Subunit interactions .....	421
Functional analyses of motifs .....	421
DNA translocation .....	421
Assembly and Its Implications.....	423
<b>BIOLOGY OF TYPE I R-M SYSTEMS .....</b>	<b>423</b>
Host-Controlled Modulation of Restriction Activity.....	423
Self-protection .....	423
Proteolytic control.....	423
Significance .....	424
Mechanisms by Which Plasmids and Phages Avoid Restriction .....	425
Detection, Distribution, and Diversity .....	426
Evolution .....	427
Relevance to Bacteria.....	429
<b>CONCLUSION AND FUTURE DIRECTIONS.....</b>	<b>430</b>
<b>ACKNOWLEDGMENTS .....</b>	<b>430</b>
<b>REFERENCES .....</b>	<b>430</b>

\* Mailing address: Institute of Cell and Molecular Biology, University of Edinburgh, Darwin Building, Mayfield Rd., King's Buildings, Edinburgh EH9 3JR, United Kingdom. Phone: 44 131 650 5374. Fax: 44 131 650 8650. E-mail: Noreen.Murray@ed.ac.uk.

## GENERAL INTRODUCTION

### Host-Controlled Modification—A Historical Perspective

Awareness of the phenomenon of restriction and modification and the consequent revolution in molecular biology grew from the observations of microbiologists in the early 1950s that the host range of a bacterial virus (phage) could be influenced by the bacterial strain in which the phage was last propagated. While phages produced in one strain of a bacterial species would readily infect a culture of the same strain, they might only rarely achieve successful infection of cells from a different strain of the same species. This finding implied that the phages carried an "imprint" that identified their immediate provenance. The occasional successful infection of a different strain resulted in the production of phages that had lost their previous imprint and acquired a new one, i.e., they acquired a new host range, hence the term host-controlled modification (see reference 5 for an early review).

In the 1960s, elegant molecular experiments showed the imprint to be a DNA modification that was lost when the phage DNA replicated within a different bacterial strain; phages that conserved one of their original DNA strands retained the modification, whereas phages containing two strands of newly synthesized DNA did not (6). The modification was shown to protect the DNA against an endonuclease, the barrier that prevented, or "restricted," the successful propagation of an incoming phage genome. Later it was proven that the modification and restriction enzymes both recognize the same target, a specific nucleotide sequence. The modification enzyme is a DNA methyltransferase (161) that methylates specific bases within the target sequence, and in the absence of the specific methylation the target sequence renders the DNA sensitive to the restriction endonuclease. When DNA lacking the appropriate modification imprint enters a restriction-proficient cell, it is therefore recognized as foreign and degraded by the endonuclease. For each unmodified target sequence, there is only a low probability that it will become modified and escape attack by the endonuclease. Since the host-controlled barrier to successful infection by phages that lack the correct modification was referred to as restriction, the relevant endonucleases have acquired the colloquial name of restriction enzymes. Similarly, the methyltransferases are more commonly termed modification enzymes. Classically, a restriction enzyme is accompanied by its cognate modification enzyme, and the two comprise a restriction and modification (R-M) system. Most restriction systems conform to this classical pattern. There are, however, some restriction endonucleases that attack DNA only when their target sequence is modified; such modification-dependent restriction enzymes do not, therefore, coexist with a cognate modification enzyme.

The classical R-M systems and the modification-dependent restriction enzymes share the potential to attack DNA derived from different strains and thereby restrict DNA transfer. They differ in that in one case an associated modification enzyme is required to protect DNA from attack by the cognate restriction enzyme, while in the other a modification enzyme specified by one strain imparts a signal that provokes the degradative activity of a restriction endonuclease in another.

Early examples of host-controlled modification, though they were not always recognized as such, were reviewed by Luria (107). Two papers in particular stimulated interest in the phenomenon. In one, Bertani and Weigle (17), using temperate phages ( $\lambda$  and P2), identified the classical R-M systems of *Escherichia coli* K-12 and *E. coli* B. In the other, Luria and

Human (108) identified a restriction system of the second, nonclassical kind.

T-even phages were used by Luria and Human as test phages, and after their growth in a mutant *E. coli* host, these phages were restricted by wild-type *E. coli* K-12 but not by *Shigella dysenteriae*. An understanding of this restriction phenomenon requires knowledge of the special nature of the DNA of T-even phages. When the DNA of T-even phages is replicated, the unusual base 5-hydroxymethylcytosine (HMC) substitutes for cytosine, and the hydroxymethyl group is then glucosylated in a phage-specific pattern. In the mutant strain of *E. coli* used by Luria and Human, glucosylation fails. The vulnerable nucleotide sequences of T4, normally protected by glucosylation, are recognized by an endonuclease in *E. coli* K-12 because they include the modified base HMC rather than cytosine residues. *S. dysenteriae* does not restrict the nonglucosylated phage because it lacks the relevant endonuclease. Later it was discovered that methylated cytosine residues in the context of the correct nucleotide sequence also evoke restriction by these modification-dependent endonucleases (145).

The biological tests for a classical R-M system are illustrated by the pioneering experiments of Bertani and Weigle (17). Phage  $\lambda$  grown on *E. coli* strain C ( $\lambda$ .C or  $\lambda$ .0), where *E. coli* C is a strain that apparently lacks an R-M system, forms plaques with poor efficiency (efficiency of plating [EOP] of  $2 \times 10^{-4}$ ) on *E. coli* K-12 because the phage DNA is attacked by a restriction endonuclease. Phage  $\lambda$  grown on *E. coli* K-12 ( $\lambda$ .K) forms plaques with equal efficiency on *E. coli* K-12 and *E. coli* C, since it has the modification required to protect against the restriction system of *E. coli* K-12 and *E. coli* C has no restriction system. In contrast,  $\lambda$ .K will form plaques with very low efficiency on a third strain, *E. coli* B, since *E. coli* B has an R-M system with different sequence specificity from that of *E. coli* K-12.

The classical restriction endonucleases of *E. coli* K-12 and B were not only the first to be detected but the first to be purified (104, 117). The demonstration that the restriction endonuclease from *E. coli* K-12 produced a digest of phage  $\lambda$  DNA comprising large DNA fragments (117) was exciting, for it implied a highly specific target for the endonuclease, a view supported by genetic studies which showed that target sequences in phage genomes could be mutated and mapped (54, 93, 124). The rational expectation was that type I restriction enzymes would cut DNA close to their target sequences, but Horiuchi and Zinder (68) showed otherwise. The enzyme from *E. coli* B cut the DNA of phage  $\phi$ 1 nonspecifically at considerable distances from the unmodified target sequences. Type I restriction enzymes therefore failed to provide the anticipated analytical reagents, but they raised the alert so that Ham Smith immediately appreciated the significance of his observation that *Haemophilus influenzae* strain Rd degraded P22 phage DNA, and as a consequence, he purified *Hind*II, the first type II restriction enzyme (160). The phenomenon of restriction, identified for type I and methylation-dependent systems, laid the foundations for modern molecular biology, yet the molecular complexity and biological importance of these systems remain to be fully understood.

### Nomenclature and Classification of R-M Systems

**Nomenclature.** Classical R-M systems are designated by a three-letter acronym derived from the name of the organism in which they occur. The first letter comes from the genus, and the second and third letters come from the species. The strain designation, if any, follows the acronym. Different systems in the same organism are distinguished by roman numerals. Thus,

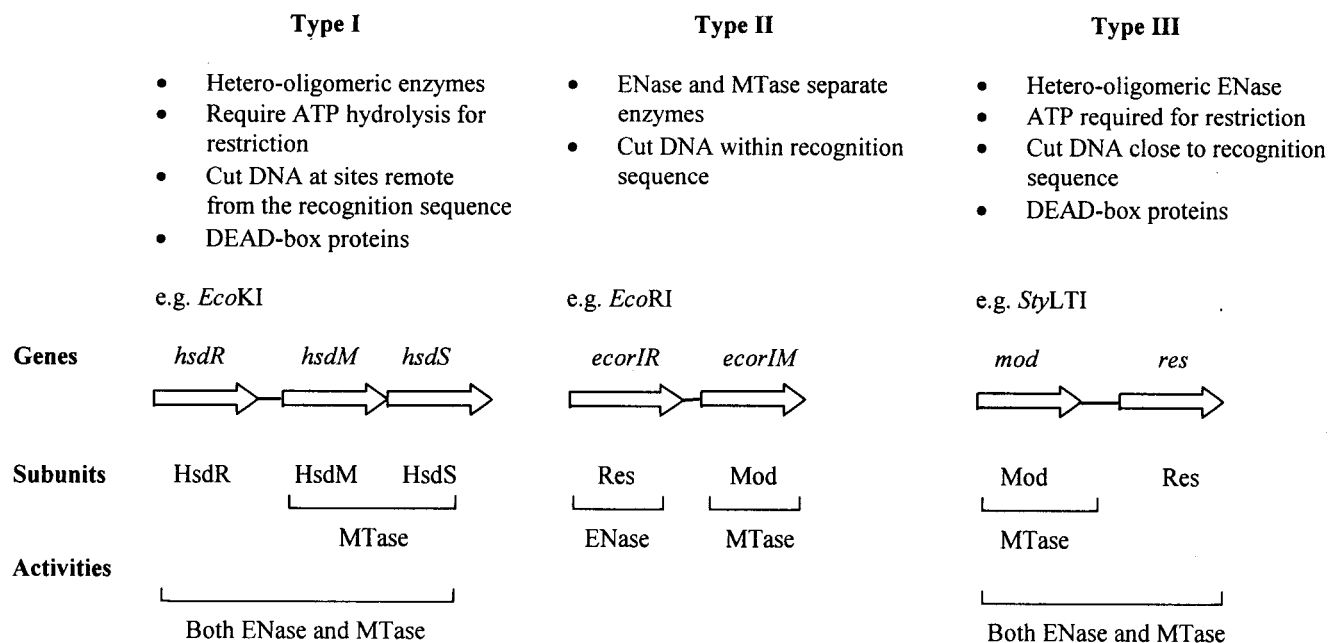


FIG. 1. Distinguishing characteristics and organization of the genetic determinants and subunits of the different types of R-M systems. ENase, restriction endonuclease; Mtase, methyltransferase. Modified with permission from a figure in reference 83.

*EcoKI* and *EcoBI* are the first classical restriction enzymes identified in *E. coli* K-12 and *E. coli* B, respectively (they are referred to as *EcoK* and *EcoB* in early papers). Restriction endonucleases and modification methyltransferases are sometimes distinguished by the prefixes R and M, respectively, but the prefix is commonly omitted, particularly for type I R-M systems.

**Classification.** R-M systems are classified on the basis of their composition and cofactor requirements, the nature of their target sequence, and the position of the site of DNA cleavage with respect to the target sequence. Currently three distinct, well-characterized types of classical R-M systems are known (I, II, and III), although a few do not fit well into any of these. The first R-M systems to be identified, those characteristic of *E. coli* K-12 and *E. coli* B, were designated type I, but the type II systems became better known, since they, unlike type I restriction enzymes, cut DNA into discrete fragments. A summary of the properties of the principal types of classical R-M systems is given in Fig. 1, and the reviews selected (5, 10, 18, 65, 83, 92, 118, 144, 149, 170, 192) include some of particular historical significance.

## INTRODUCTION TO TYPE I R-M SYSTEMS

### General Characteristics

Type I R-M systems are multifunctional enzymes that can catalyze both restriction and modification. S-Adenosylmethionine (AdoMet) is the cofactor and methyl donor for the methyltransferase activity; the endonucleolytic activity requires ATP, AdoMet, and  $Mg^{2+}$ . The nucleotide sequences recognized by type I enzymes are asymmetric and comprise two components, one of 3 or 4 bp and the other of 4 or 5 bp, separated by a nonspecific spacer of 6 to 8 bp. All known type I enzymes methylate adenine residues, one in each component of the target sequence, but on opposite strands. The type I R-M enzyme binds to its target sequence, and its activity as an endonuclease or a methyltransferase is determined by the

methylation state of the target sequence. If the target sequence is unmodified, the enzyme, while bound to its target site, is believed to translocate, or pull, the DNA towards itself simultaneously in both directions in an ATP-dependent manner. This translocation process brings together enzymes bound to different sites on the same molecule, and it is thought that DNA cleavage occurs when translocation is impeded, either by collision with another translocating complex or by the topology of the DNA substrate.

A type I restriction enzyme comprises three subunits encoded by three closely linked genes, *hsdR*, *hsdM*, and *hsdS*. The acronym *hsd* was chosen at a time when R-M systems were referred to as host specificity systems and *hsd* denotes "host specificity of DNA." *hsdM* and *hsdS* are transcribed from the same promoter, but *hsdR* is transcribed from its own promoter. The two subunits encoded by *hsdM* and *hsdS*, HsdM and HsdS (often referred to as M and S), are both necessary and sufficient for methyltransferase activity. The third subunit (HsdR or R) is required for restriction. The S (specificity) subunit includes two target recognition domains (TRDs) that impart target sequence specificity to both the restriction and modification activities of the complex. HsdM includes the binding site for AdoMet and the active site for DNA methylation; HsdR includes the active site for ATP hydrolysis and other sequences essential for DNA translocation and endonuclease activity.

### The Family Concept

The finding that type I R-M systems exist as closely related members of a family has been of fundamental value to their analysis (10). Evidence for related systems was first indicated by the demonstration that mutants with defects in the allelic genes encoding *EcoKI* and *EcoBI* could complement each other. On the basis of such tests, it was inferred that each enzyme comprised three subunits, that the subunits of *EcoKI* and *EcoBI* were interchangeable, and that the subunit encoded by one gene, *hsdS*, confers target sequence specificity on the multimeric complex (22, 58, 69). *EcoKI* and *EcoBI* became the

founder members of a family of type I systems, type IA. The essential difference between two members of one family resides in the regions of the HsdS subunit that confer sequence specificity.

Hybridization screens of bacterial DNAs and serological screens of bacterial extracts first suggested that allelic genes might also encode sufficiently dissimilar type I R-M systems to warrant their separation into different families (122). As expected, the nucleotide sequences of the *hsd* genes for *EcoKI* and *EcoBI* would hybridize to each other and antibodies raised against *EcoKI* reacted with *EcoBI*, but in contrast, DNA probes comprising the *EcoKI* genes failed to hybridize with those of *E. coli* 15T<sup>-</sup>, which encoded *EcoAI*; similarly, antibodies against *EcoKI* did not cross-react with *EcoAI*. The *hsd* genes in these two strains behave as alleles in genetic tests (7) but have very different nucleotide sequences (34, 80). *EcoAI* defines a second family of type I systems, type IB.

A third family, type IC, headed by *EcoR124I*, includes plasmid-encoded members (141) and a chromosomally encoded relative (*EcoprrI*) identified in a natural isolate of *E. coli*. The first representative of a fourth family, ID (180), is the R-M system of *Salmonella enterica* serovar *blegdam*, identified initially on the basis of biological tests (26). The genes encoding the ID systems, but not those for *EcoprrI* (183), map to the same region of the *E. coli* chromosome as those for type IA and type IB. Currently, each type I R-M system identified in *E. coli*, or in a close relative, has been allocated to one of the four families (the members are listed later in Table 2).

## ENZYMES

### Introduction

The restriction endonucleases from three families of type I systems (IA, IB, and IC) have been purified and characterized. Each is a large oligomeric complex of relative molecular weight ~400,000 to 500,000, which in the presence of Mg<sup>2+</sup>, ATP, and AdoMet functions as an endonuclease on a DNA substrate that includes unmethylated target sequences but catalyzes the transfer of methyl groups from AdoMet to DNA substrates that include hemimethylated target sequences (187). Early evidence indicated that *EcoKI* comprises subunits of three sizes, with approximate molecular weights of 135,000, 62,000, and 52,000. Estimates of the relative amounts of each subunit in *EcoKI* indicated two of each of the larger polypeptides and one of the smallest (118). An analysis of the polypeptides specified by the cloned *hsd* genes of *E. coli* K-12 permitted the correlation of polypeptides with genes (151) and the consequent suggestion that the stoichiometry of the subunits within the complex is R<sub>2</sub>M<sub>2</sub>S<sub>1</sub>.

Convincing evidence that endonuclease activity is associated only with the R<sub>2</sub>M<sub>2</sub>S<sub>1</sub> complex was obtained much later (45). For a time it appeared that *EcoR124I* and -II, members of the IC family, have one rather than two R subunits and that they retain endonuclease activity in the absence of AdoMet (75). However, it has now been shown that the *EcoR124I* complex has a tendency to lose one HsdR polypeptide (76) and that AdoMet copurifies with *EcoR124I* and other type IC enzymes (41, 133; P. Janscak and T. A. Bickle, personal communication). Type IB complexes readily lose both HsdR subunits (167), but it seems probable that the predominant active complex for any type I R-M system includes two HsdR subunits.

A modification enzyme without endonuclease activity was demonstrated in extracts from *E. coli* B, in addition to the larger complex with both activities (53, 100). Strains of bacteria encoding *EcoKI* (169), *EcoAI* (167), and *EcoR124I* (174) are

now known to possess a complex with only methyltransferase activity, as well as the large R-M complex. In all cases, the stoichiometry of the enzyme lacking endonuclease activity is M<sub>2</sub>S<sub>1</sub>.

In the following sections, evidence for the functional domains within each subunit is summarized before the activities of the enzyme complexes are considered.

### Specificity Subunit—HsdS

**Variable sequences.** When the sequences of specificity genes (*hsdS*) of members of the type IA family were searched for differences that would correlate with the recognition of different target sequences, each pairwise comparison revealed two long regions (~450 to 500 bp) of apparently unrelated sequence (referred to as variable regions), in addition to some minor differences in the intervening conserved sequences. Together, the two variable regions comprise the major portion of the gene, but given the bipartite nature of the target sequences, it was inescapable that each variable region might encode polypeptide sequences responsible for the recognition of one component of the target sequence (62). Subsequently, comparisons of the nucleotide sequences of *hsdS* genes within the IB (80), IC (64), and ID (A. J. B. Titheradge and N. E. Murray, unpublished data) families also revealed two long variable regions flanking a conserved region, as shown for a IC member in Fig. 2a.

Analyses of hybrid *hsdS* genes provide support for the notion that the variable regions encode TRDs. It was demonstrated for the IA family that the specificity conferred by a chimeric HsdS polypeptide has one half of its recognition sequence in common with each parental target sequence (55, 127), a finding confirmed for members of the IB (179) and IC (64) families. It was demonstrated for the first hybrid *hsdS* gene that any amino acid differences in the central conserved region are irrelevant to sequence specificity and that when two TRDs from different families recognize the same nucleotide sequence, they have significant similarity, e.g., 50% identity, despite the absence of much similarity between the sequences of the remainder of HsdS or of HsdM and HsdR (33).

**Conserved sequences.** Two structural roles are anticipated for the conserved regions of HsdS polypeptides. These are the maintenance of the relative positions of the two TRDs and the specific associations of HsdS with other components of the R-M complex. Evidence that the conserved peptide sequence that links the two TRDs serves to position the TRDs on the target sequence came from the chance occurrence of a derivative of *EcoR124I* with a new specificity. The sequence of the *hsdS* gene of this R-M system, *EcoR124II* (originally referred to as *EcoR124/3*), is consistent with a recombination event in which unequal crossing-over between a misaligned 12-bp duplication in the central conserved sequence of the *hsdS* gene, or slipped mispairing during DNA replication, has led to a triplication of that sequence. As a result, the two components of the target sequence of *EcoR124II* are separated by 7 bp rather than the 6 bp in the target sequence of *EcoR124I* (63, 142). The region in which the four additional amino acids occur is predicted to be strongly helical, and the increase in length is likely to be sufficient to accommodate a change in spacing of 1 bp (142).

**Imposition of symmetry.** In addition to the regions conserved in HsdS subunits of the same family, there are regions of similarity within each HsdS subunit (Fig. 2a). For *EcoKI* (type IA), this was noted by Argos (8) for regions in the central and carboxy-terminal conserved sequences. In the IB family, a segment of amino acids is common to both the amino-terminal



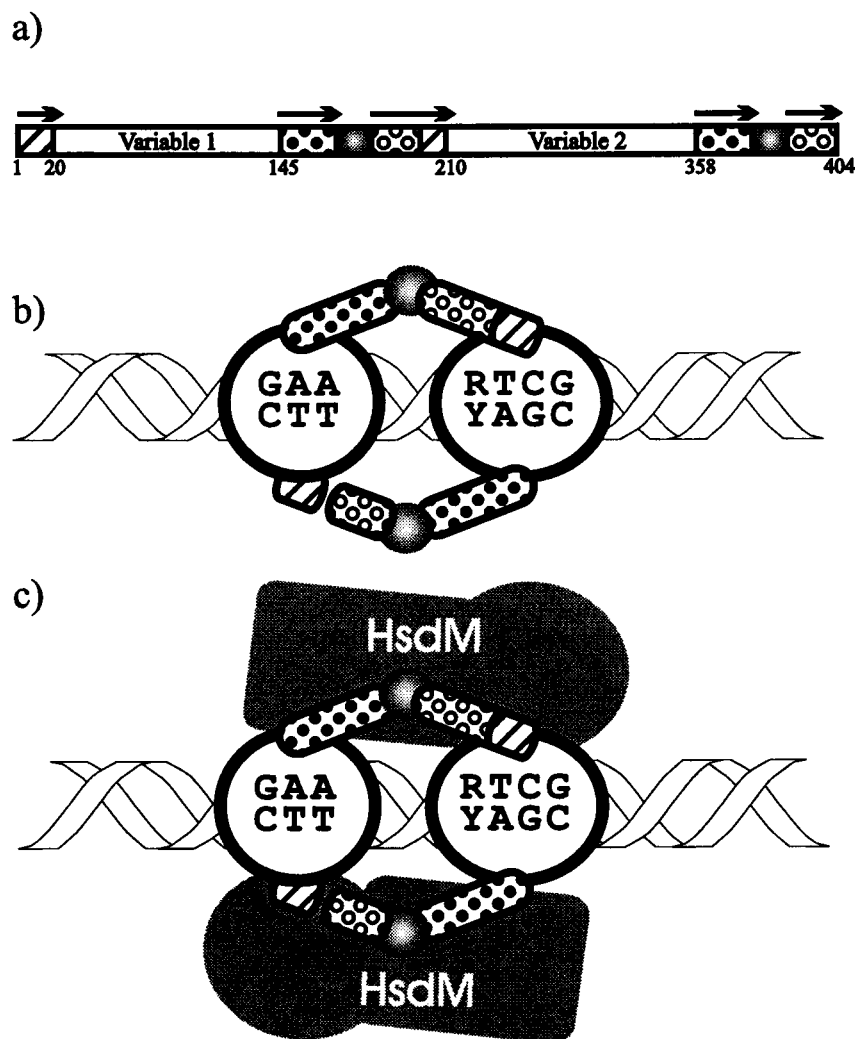


FIG. 2. Specificity polypeptide of *EcoR124I*. (a) Organization indicating the two variable regions (TRDs) and the regions conserved in sequence for all members of one family. Repeats of similar sequence (below the arrows) are identified by their conserved marking. The number of the amino acid residue at the beginning and end of each region is given below the diagram. (b) Model of Kneale (86), in which the repeated sequences form linkers joining the TRDs in a rotationally symmetrical configuration. The nucleotide sequences identified by each TRD are shown. (c) Consequent model of the methyltransferase, in which the two HsdM subunits bind to the linker region to generate an enzyme with pseudodyad symmetry.

and central conserved sequences (Fig. 3c), in addition to regions similar to those in *EcoKI* (80). The analysis of sequence similarities in the specificity subunit of a type IC member indicated that a sequence in the central conserved region was incompletely repeated at the carboxy terminus, with the remainder of the repeat located at the amino terminus (Fig. 2a). This "split repeat" led Kneale (86) to propose a model in which the amino and carboxy termini of HsdS are in close proximity so that they associate to form a linker domain similar to that provided by the central conserved sequence (Fig. 2b). These two domains of similar sequence would then dictate the symmetrical association of the two HsdM subunits (Fig. 2c). The repeated sequences identified within the HsdS subunits of the type I systems of the gram-positive bacterium *Lactococcus lactis* (153) are remarkably reminiscent of those identified in the type IC family, and a close analysis of members of the type IA family finds some evidence for a short sequence at the amino terminus of HsdS that has similarity with part of the repeat in the central conserved region (86).

The symmetrical organization of the HsdM subunits was

implicit in an earlier model (27). The present model (Fig. 2c) and that of Willcock et al. (191) indicate a methyltransferase structure with twofold rotational symmetry in which inversely oriented TRDs will each make contact with one HsdM subunit (113).

This symmetrical configuration of the domains within HsdS is supported by the following observations. (i) Truncated forms of the HsdS subunit of either *EcoDXXI* or *EcoR124I*, in which the carboxy half of the polypeptide is missing but the central conserved sequence is retained, associate to form an active enzyme that recognizes a bipartite target sequence that is a palindromic version of the trinucleotide specified by the amino-terminal TRD (1, 113). Hence, *EcoDXXI* recognizes TCA(N<sub>7</sub>)RTTC, while the derivative with a truncated HsdS polypeptide recognises TCA(N<sub>8</sub>)TGA (Fig. 3a). (ii) The amino-terminal TRD of *SlySKI* and the carboxy-terminal TRD of *EcoR124I* are very similar (36% identity) despite being members of different families. This similarity correlates with a similarity of target sequence evident in the complementary strands (Fig. 3b), a finding consistent with the amino- and carboxy-

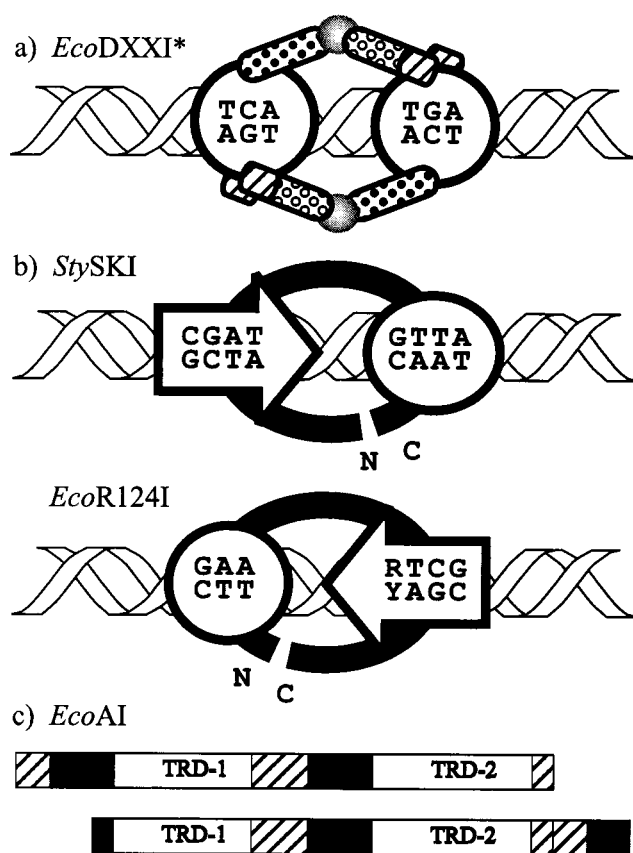


FIG. 3. Experimental support for the symmetrical arrangement of the TRDs (large open circles or arrows) within specificity subunits. (a) Predicted arrangement of two truncated subunits of *EcoDXXI* and the consequent recognition of a hyphenated symmetrical target sequence. (b) The HsdS subunits of *StySKI* (type IB) and *EcoR124I* (type IC). The IB family of enzymes have a long N-terminal conserved sequence, while the IC family has a long C-terminal conserved tail. The amino-terminal TRD of *StySKI* and the carboxy-terminal TRD of *EcoR124I* (indicated by arrows) have 36% amino acid identity. The DNA targets recognized are indicated. The complement of the *EcoR124I* 3' target sequence (CGAT) is a degenerate version of the *StySKI* 5' target sequence (CGAT), consistent with a symmetrical organization of two similar TRDs. (c) Circularly permuted variants of the specificity gene of *EcoAI* (type IB) can retain activity. The normal organization of regions within an HsdS polypeptide is given in the first diagram, followed by an active but circularly permuted variant. The black segments identify the well-conserved repeats.

terminal TRDs being inversely oriented when they bind their target sequences (179). (iii) Permutations of the *hsdS* gene of *EcoAI* have been made in which HsdS remained active when a sequence from the N-terminal conserved region was transferred to the carboxy terminus (Fig. 3c) (74).

**Association of subunits.** Interaction with other subunits was an anticipated role for the conserved sequences of HsdS polypeptides (62). Convincing evidence in support of this prediction derives from the truncated derivatives of HsdS polypeptides of *EcoDXXI* (Fig. 3a) and *EcoR124I*. Two truncated polypeptides can substitute for one normal HsdS subunit if the subunits retain their ability to interact with HsdM. Active complexes, i.e., those in which HsdS retains the ability to bind HsdM, are recognized by their new specificities (1, 109, 113). Analyses of truncated polypeptides implicate the three conserved regions of HsdS (Fig. 2a) in binding HsdM (109). As predicted, a deletion within the central conserved region of an HsdS subunit prevents binding to HsdM (189).

Point mutations have given some indication of the regions of

HsdS that are involved in protein-protein interaction. One mutation in the *hsdS* gene of *EcoKI* destabilized the methyltransferase at high temperature, consistent with a defect in the binding of HsdM to HsdS (197; D. T. F. Dryden, V. Zinkevich, and K. Firman, personal communication). Rare mutations in the *hsdS* genes of both *EcoKI* (131) and *EcoR124I* (189) lead to a restriction-deficient, modification-proficient ( $r^{-}m^{+}$ ) phenotype, the phenotype predicted if HsdR can no longer associate with the methyltransferase. These mutations may occur preferentially at the borders of conserved and variable regions. Weiserova and Firman (189) suggested that the mutations identify regions of importance for the assembly of the R-M complex, though not necessarily sites of interaction between HsdS and HsdR.

**DNA specificity.** The HsdS subunits of type I R-M enzymes do not possess obvious DNA-binding motifs within their TRDs of 150 to 180 amino acids. No three-dimensional structural analysis of a type I R-M enzyme or any component of a type I R-M enzyme has been achieved. There is, therefore, no direct evidence to identify which amino acids within a TRD interact with DNA. Two approaches to understanding the mechanism of DNA recognition have been made: one relies on modeling, and the other relies on random mutagenesis of a TRD. Fortunately, the strongest predictions from modeling were obtained for the amino-terminal TRD of *EcoKI*, the TRD subjected to mutagenesis.

**Molecular model.** Sturrock and Dryden (166) supplemented sequence data from known type I R-M systems with data for putative systems identified in genomic sequences to derive a molecular model for the recognition region of a TRD. These authors used amino acid sequences combined with secondary-structure prediction to align 51 TRDs. The inclusion of secondary-structure prediction enhances the strength of the amino acid alignments, making distant similarities more apparent. This is particularly helpful because the amino acid identities between TRDs that recognize different target sequences are usually less than 25%. The alignments suggest a common tertiary structure, and secondary-structure predictions with strong similarity to the known structure of the TRD of the type II *HhaI* methyltransferase. Of the 51 sequences compared, that of the amino-terminal TRD of HsdS from *EcoKI* shows the closest similarity to the TRD of *M. HhaI*, sufficiently so to suggest that *EcoKI*, like *M. HhaI*, might interact with DNA via two short polypeptide loops flanking a  $\beta$ -strand.

The experimental approach of O'Neill et al. (131) aimed to localize the protein-DNA interface by random mutagenesis. It was anticipated that amino acids that could be changed without loss of R-M activity were unlikely to be involved in target recognition, while substitutions that resulted in an  $r^{-}m^{-}$  phenotype would include amino acids involved in a specific interaction with DNA. Most of 101 substitutions affecting 79 of the 150 residues, including quite severe changes, had no detectable effect on phenotype; changes at only seven positions conferred an  $r^{-}m^{-}$  phenotype. Five of the seven residues identified are in an interval between residues 80 and 110 which includes the predicted loop- $\beta$ -strand-loop: the model places two of these (residues 91 and 107) close to the protein-DNA interface (Fig. 4a and b). Three further residues (92, 95, and 103), all close to the DNA in the model, were changed by site-directed mutagenesis, and substitutions for each impaired both modification and restriction. Additional residues within the predicted  $\beta$ -strand and second loop have been changed (M. O'Neill and N. E. Murray, unpublished data); a mutant with a substitution for residue 105 is  $r^{-}m^{-}$ , while those with substitutions for residues 94 and 106 retain only modification activity ( $r^{-}m^{+}$ ).

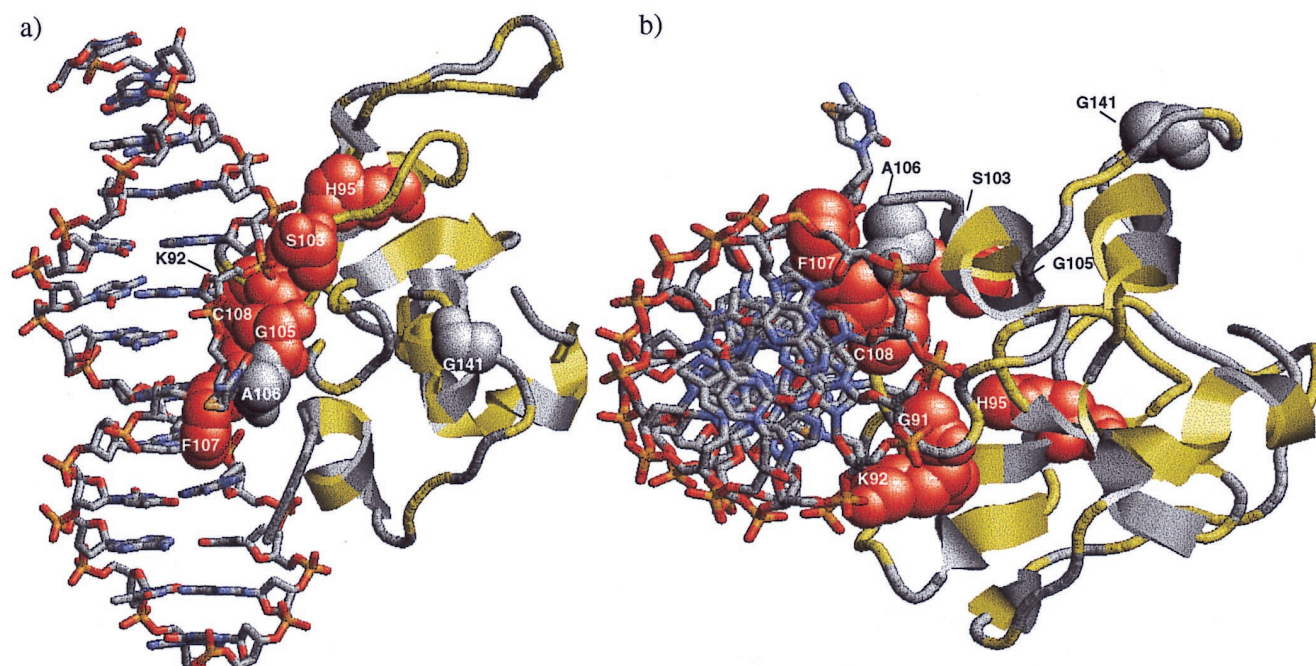


FIG. 4. Model of the amino acid segment comprising residues 43 to 157 of the amino terminus of *EcoKI* interacting with DNA as proposed by Sturrock and Dryden (166). (a) Side view. (b) Bird's eye view. Residues in red are substitutions that cause the loss of restriction and modification; those in yellow have no detectable effect on activity. Residue 141, in gray, was previously thought to lead to a loss of both activities but is now known to impair rather than inactivate. The DNA structure is from the complex of *M. HhaI* with its DNA target sequence (85) and therefore shows an extrahelical cytosine rather than adenine. The figure is modified with permission from that published in O'Neill et al. (131).

Two substitutions (residues 57 and 141) outside the predicted loop- $\beta$ -strand-loop region have been reported to confer an  $r^{-}m^{-}$  phenotype (131), but neither residue appears to be relevant to specificity. Only one of three substitutions identified for residue 57 conferred an  $r^{-}m^{-}$  phenotype, and it is now known that this phenotype results from a mutation in *hsdM*, not the change at position 57 (M. O'Neill and N. E. Murray, unpublished data). For residue 141, additional tests indicate that the G-to-A substitution severely impairs but does not destroy the restriction and modification activities. A tyrosine residue previously shown by cross-linking to be in close proximity to the DNA (28) is also outside the loop- $\beta$ -strand-loop region. This tyrosine residue has been replaced by cysteine, and the mutant retains modification activity (M. O'Neill and N. E. Murray, unpublished data). Currently, the genetic data and the analysis of mutant *EcoKI* enzymes correlate well with the predictions of the structural model for the TRD.

Two lines of evidence from other systems are also consistent with the structural model. First, Taylor et al. (176) used chemical modification to identify accessible lysine residues in *EcoR124I*. Three lysine residues in the carboxy-terminal TRD were especially susceptible to modification in the absence of bound DNA. The most strongly modified residue lies within the second of the proposed loops (166). Second, the three residues identified by chemical modification of *EcoR124I* are conserved in the amino-terminal TRD of the type IB system, *SlySKI* (179), which recognizes the same target sequence as the carboxy-terminal TRD of *EcoR124I*.

**HsdM needed for DNA binding.** The HsdS subunits of *EcoKI* (193) and *EcoR124I* (132) are insoluble in the absence of HsdM, but the HsdS subunit of *EcoR124I* has been produced in soluble form as a glutathione-S-transferase (GST) fusion product (97). The resulting fusion protein is unable to

make specific complexes with DNA. The GST moiety at the N terminus interferes with the recruitment of the second HsdM subunit, preventing assembly of the active methylase; its removal allows HsdS to assemble with HsdM units to form an active modification enzyme (115). These findings are consistent with the model (Fig. 2c) in which the conserved sequence at the N terminus forms part of the region that contacts HsdM. The HsdS subunit of *EcoAI*, a type IB enzyme, is soluble in the absence of HsdM, but it, too, requires HsdM to make sequence-specific complexes (74). For *EcoKI*, even the association with one HsdM subunit is sufficient to promote a sequence-specific interaction, though with a higher  $K_d$  than the full  $M_2S_1$  complex (136). It would appear that HsdM is important for the positioning of the domains of HsdS, at least in part by maintaining its rotationally symmetrical configuration (74, 86).

#### Modification Subunit—HsdM

Like HsdS, HsdM is essential for restriction as well as modification. AdoMet, the methyl donor for modification, is an essential cofactor for restriction (117), and early experiments showed that AdoMet binds to the HsdM subunit of *EcoKI* (24). Consistent with this, the amino acid sequence predicted from the nucleotide sequence of the *hsdM* gene of *E. coli* K-12 (106) includes a version of a motif, N/DPPF/Y/W, that is characteristic of both N-6 adenine and C-4 cytosine methyltransferases and is now generally referred to as motif IV (for reviews of methyltransferases, see references 29, 42, and 129).

**Active site.** The HsdM subunit is well conserved (~90%) within a family (156, 183), but comparisons between members of different families generally indicate only 25 to 30% identity (156). Site-directed mutagenesis of the *hsdM* gene of *E. coli* K-12 demonstrated the relevance of two conserved sequences



to methyltransferase activity (191). A change in the sequence predicted to be motif I (D/E/SXFXGXG) abolished the binding of AdoMet, while changes in motif IV (N/DPPF/Y/W) prevented catalysis but did not affect binding of the cofactor. Nevertheless, a tryptophan residue substituted for phenylalanine in motif IV of *EcoKI* is sufficiently close to the AdoMet-binding site to enhance cross-linking with the methyl donor (191). Subsequently, molecular modeling of type I HsdM subunits, based on sequence alignment and predicted secondary structures, suggested a domain in HsdM subunits which resembles that of the  $\gamma$  class of type II N-6 adenine methyltransferases (44) and includes the six motifs found in the catalytic domain of *M·TaqI* (99). The model of the catalytic domain for the HsdM subunit of *EcoKI* is consistent with the location of proteolysis-sensitive sites (31) and the mutational analysis of Willcock et al. (191). In this model, the glycine at amino acid residue 177 (motif I) is located close to the cofactor, and the phenylalanine at residue 269 (motif IV) is positioned at the edge of the active site, where it can interact with the target adenine should this be flipped out of the DNA helix, as has been shown for the cytosine residue during methylation by *M·HhaI* (85). Base flipping allows access to the target base and is therefore predicted to be common to all DNA methyltransferases as well as some DNA repair enzymes (150).

### Modification Enzyme

A complex with only methyltransferase activity has been purified from representatives of the type IA, IB, and IC families. The stoichiometry of this active methyltransferase is  $M_2S_1$  (167, 174). *EcoKI* and *EcoBI* (type IA) dissociate to  $M_1S_1$  and  $M_1$  (43, 45, 100, 136), but for *EcoKI* the active form has been shown to be  $M_2S_1$  (43). The type I modification enzymes catalyze the transfer of the methyl group from AdoMet to the N-6 position of specific adenine residues in their respective target sequences, probably utilizing the cationic- $\pi$  interactions proposed for those type II enzymes that methylate adenine at the N-6 position or cytosine at the N-4 position (152).

**Discrimination of methylation state.** The three type I families differ significantly in their relative responses to hemimethylated versus unmethylated target sequences. *EcoAI*, the type IB representative, like most known methyltransferases, shows little or no preference for hemimethylated DNA (167). In contrast, in vitro experiments indicate that members of the IA and IC families have a very strong preference for hemimethylated DNA (approximately 100-fold compared to unmethylated DNA); the data for the type IA system have been obtained from experiments with both plasmid DNA (169) and oligonucleotide duplexes (43) as substrates, and those for IC were obtained from oligonucleotide duplexes (175). Consistent with the in vitro evidence, unmodified phage DNA is a poor substrate in vivo for methylation by a type IA modification system (82, 105), but in apparent contradiction to the in vitro evidence, unmodified phage DNA was found to be a good substrate in vivo for methylation by an  $r^-m^+$  type IC system (S. Makovets and N. E. Murray, unpublished data).

Mutants of potential relevance to the discrimination between hemimethylated and unmethylated target sequences ( $m^*$  mutants) were selected for *EcoKI* by their enhanced ability to modify unmethylated target sequences in phage  $\lambda$ . The restriction proficiency of all the mutants was impaired, though usually only slightly. Mutants selected by their  $r^-m^+$  phenotype were also found to have an  $m^*$  phenotype. Analysis of 22  $m^*$  mutants identified nine residues (14 substitutions), all within the N-terminal third of HsdM (82). Five amino acid substitutions, affecting three of these nine residues, resulted in

the absence of any detectable restriction activity in vivo. Methyltransferase has been purified from each of four  $m^*$  mutants, two  $r^-$  and two  $r^{+/-}$  (D. T. F. Dryden and N. E. Murray, unpublished data). The rate of methylation of unmethylated DNA was enhanced for all the enzymes, though to very different degrees (from 2- to 240-fold). The enzyme with the greatly enhanced rate of activity was shown to have a marked reduction in preference for hemimethylated DNA. An enhanced ability to modify unmodified target sequences could lead to competition between the two activities of the R-M complex and a consequent reduction in restriction. Competition does not entirely explain the present data, in which no direct correlation was found between the enhanced rate of methylation and the deficiency in restriction proficiency. The complete absence of restriction activity may require a more germane explanation than merely the inability to compete with the enhanced modification activity. It could be, as suggested by Kelleher et al. (82), that in these mutants an unmethylated target sequence does not trigger the enzyme to adopt the restrictive mode.

**DNA binding.** Enzymes of the IA and IC families have been investigated by gel retardation and footprinting experiments. The studies were aimed not only at understanding how the enzymes discriminate their target sequences from other DNA sequences, but also how they distinguish the methylation state of their target sequence.

DNA-binding studies for the *EcoKI* methyltransferase showed that differences in binding affinity contribute to the distinction between specific and nonspecific DNA sequences (135, 136). However, the methylation state of the recognition sequence had no effect on the binding affinity, suggesting that the preference for a hemimethylated rather than an unmethylated DNA substrate is effected mainly at the level of catalysis. Similarly for the type IC *EcoR124I* methyltransferase, an increase in  $k_{cat}$  rather than a decrease in  $K_m$  was suggested to be the reason for the faster methylation of hemimethylated DNA (175). Both *M·EcoKI* and *M·EcoR124I* cover approximately 25 to 30 bp of DNA, as determined by exonuclease III and DNase I footprinting experiments (114, 135, 136, 175). The efficiency of binding by *M·EcoKI* is enhanced by AdoMet (135), and the inactive, partially assembled form ( $M_1S_1$ ) covers the same length of DNA as the active methyltransferase ( $M_2S_1$ ), indicating that the two subunits of HsdM are located on either side of HsdS, away from the helical axis of the DNA (136).

Interference footprinting experiments show that *M·EcoKI* (134) and *M·EcoR124I* (116) make contacts in the major groove of the DNA helix. Only in the case of the IA family has it been possible to examine the effect of the cofactor AdoMet in any detail, given that AdoMet copurifies with IB and IC enzymes. The presence of AdoMet has a striking effect on the interference pattern for *M·EcoKI* bound to unmodified DNA but not when bound to either hemimethylated or fully methylated DNA. For *M·EcoKI*, the methylation state of the target sequence therefore affects the conformation of the protein at the DNA interface, and it would appear that AdoMet could play an important role in the discrimination between unmodified and modified DNA (134).

**Base flipping.** The footprinting of *M·EcoR124I* provides the following circumstantial evidence that the adenine residues flip out to provide access for methylation (114). Two sites that are hypersensitive to hydroxyl radical cleavage have been identified within the target sequence, one per strand, each associated with the adenine that is the substrate for methylation. The authors argue that it is unlikely to be fortuitous that the hypersensitive site on each strand coincides with the adenine that is the target for methylation. A plausible explanation is that bind-



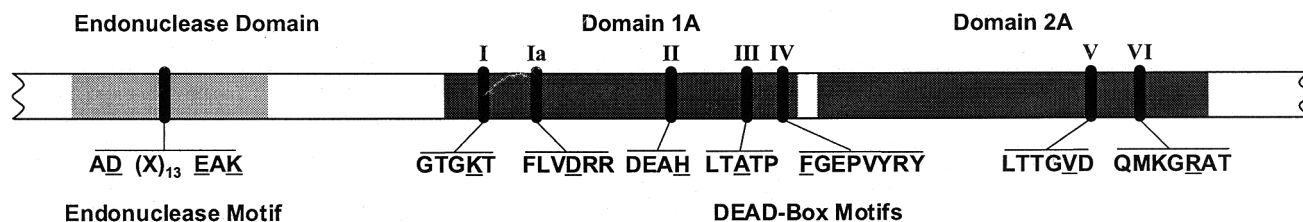


FIG. 5. Domains and motifs of HsdR of *EcoKI*. The N- and C-terminal regions are omitted, since their roles are not known. The two domains that include the DEAD-box motifs correlate with 1A and 2A, as determined for the structures of DNA helicases (36). Substitutions for the underlined amino acids confer a restriction-deficient phenotype.

ing of the enzyme induces a marked conformational change in the structure of the sugar-phosphate backbone of the DNA in the region around those bases that are the targets for methylation. Mernagh et al. (116) also showed that M · *EcoR124I* binds more strongly when either uracil or an abasic site is substituted for one of the target adenines. Again, this would be consistent with the idea that the adenine residues, like the cytosine residues in the *HhaI* target sequence (85, 150), are flipped out during the methylation reaction. Base analog experiments with the *HhaI* methyltransferase have shown that this enzyme binds more strongly to substrates in which the target base is mismatched (84).

#### Restriction Subunit—HsdR

The polypeptide sequences of HsdR subunits representative of the four families of type I systems have only 20 to 30% identity (180). All these sequences, however, include motifs characteristic of ATP-binding proteins, consistent with the ATP dependence of restriction. In addition, there are other conserved sequences indicating the presence of motifs previously identified in ATP-dependent helicases and putative helicases (59, 123, 180). There are several superfamilies of ATP-dependent helicases. Members of superfamilies 1 and 2 include the motif DEAD or a variant of this motif and are often referred to as DEAD-box proteins (103). The motifs in HsdR subunits are indicative of superfamily 2 (59). The structures of DEAD-box helicases indicate that the motifs form a nucleoside triphosphate (NTP)-binding pocket and a portion of a nucleic acid-binding site (88, 186; see reference 64a for a recent review). It is suggested that the conserved motifs define an "engine" that powers translocation on single-stranded DNA and unwinding of duplex DNA (64a). Velankar et al. (186) present an elegant model for the coupling of the energy to DNA translocation in which the enzyme "inchworms" along a single DNA strand using unpaired bases. Nearly all reported mutations affecting DEAD-box motifs impair the hydrolysis of NTP or the coupling of NTP hydrolysis to nucleotide unwinding (64a). Mutagenesis of the *hsdR* gene of *EcoKI* showed that each of the seven DEAD-box motifs is essential for restriction *in vivo* (35, 188).

The number of bacterial genomes for which sequences are available has increased significantly since the alignment of HsdR sequences reported by Titheradge et al. (180). The additional sequences have improved the reliability of alignments and the prediction of the secondary structure. Comparisons of HsdR sequences with Rep and PcrA, DNA helicases of known structure (88, 186), suggest that the HsdR subunits have the same secondary structure as the helicases in the region that includes the DEAD-box motifs (36). In Rep and PcrA, the motifs reside in two domains that couple ATP hydrolysis to DNA helicase activity (20). The fragmentation patterns produced by limited proteolysis of HsdR are consistent with the

location of the DEAD-box motifs in two domains similar to those first observed in Rep and PcrA, in which the DEAD-box motifs cluster around a cleft between two domains (36).

The earlier alignments of HsdR sequences (180) detected an additional conserved sequence in the N-terminal part of the polypeptide which proteolysis experiments indicate to be in a separate domain from those including the DEAD-box motifs (36). This additional conserved sequence (Fig. 5) has similarities with motifs associated with DNA nicking in both type II restriction enzymes and the RecB family of nucleases. Site-directed mutagenesis proved the relevance of this motif to the endonuclease activities of *EcoAI* (78) and *EcoKI* (36, 37).

#### R-M Complex

The earliest biochemical interest in restriction enzymes was as proteins that made specific interactions with DNA by a recognition process assumed to be intolerant of errors. The question of specificity for type I R-M systems extends to include the mechanism by which each enzyme not only recognizes the methylation state of its target sequence, but then reacts as a methyltransferase if the target sequence is already hemimethylated or as an endonuclease if it is unmethylated. Finally, of special interest is the mechanism by which type I R-M enzymes translocate DNA for considerable distances before breaking phosphodiester bonds in both strands of the DNA duplex. Associated with this mechanism is the problem of what halts translocation and triggers DNA cleavage. Complete answers to all of these complex molecular problems are not yet available (see reference 170 for a recent review).

**DNA binding.** The restriction pathway is presumed to be initiated by the ATP-dependent conformational change originally reported for *EcoKI* (19) but analyzed only recently by footprinting (137). *EcoKI* remains the best-studied R-M complex in terms of sequence-specific binding, the effect of the cofactor AdoMet on representatives of the type IB and IC enzymes being less easy to assess because it is difficult to separate AdoMet from these enzymes (41, 133; P. Janscak and T. A. Bickle, personal communication). *EcoKI* binds to DNA in the absence of ATP. A strong footprint of 42 to 46 bp is detected only if the DNA includes a target sequence. On the addition of ATP in the presence of AdoMet, unmethylated or hemimethylated target sequences remain protected, and the footprint shrinks to 30 bp; both fully modified target sequences and nonspecific DNA lose all protection. ATP and AdoMet are both needed for the conformational change in response to target sequences, though *S*-adenosyl homocysteine may be substituted for AdoMet and a nonhydrolyzable analog may be substituted for ATP (137).

**Recognition of methylation state.** In the presence of the functional cofactors, *EcoKI* methylates hemimethylated DNA but initiates ATP-dependent translocation if the target sequence is unmethylated. AdoMet has long been implicated in

determination of the methylation state of the target sequence. Burkhardt et al. (27) suggested that the HsdM subunits would use the methyl group of the AdoMet as a probe for the presence of the methylated base in the major groove. For modified DNA, steric hindrance from the methyl groups on the adenine residues could prevent a conformational change. Only if both adenines were unmethylated would the HsdM subunits enter the major groove to give the "closed conformation," in which the HsdR subunits are appropriately positioned to initiate DNA translocation. Recent data indicate that the TRDs of HsdS enter the major groove (28), and by analogy with type II modification enzymes, the bases that are to be modified are anticipated to flip out of the helix (116). Nevertheless, AdoMet could still serve as a probe for the methylation state of the target DNA (134). Steric hindrance by AdoMet could block the positioning of a methylated base within the active site after the adenine residues become exposed. This refinement of the model maintains a critical role for AdoMet in restriction as well as modification, and a simple prediction is that loss of AdoMet binding to HsdM should cause an  $r^{-}m^{-}$  phenotype. Consistent with this, a mutant of *EcoKI* in which AdoMet binding to the HsdM subunit is blocked as the result of a single-amino-acid substitution in motif I (D/E/SXFXGXG) (191) is deficient in restriction as well as modification (V. Doronina and N. E. Murray, unpublished data).

**Communication.** The means by which the enzyme communicates the methylation state of its target sequence has been probed by looking for mutations in *hsdM* or *hsdS* that prevent access to the restriction pathway. Analysis of an *hsdM* or *hsdS* mutant with an  $r^{-}m^{+}$  phenotype could identify an enzyme locked into the modification mode, or it could indicate that assembly of the R-M complex is prevented. The  $r^{-}m^{+}$  mutants of the *EcoKI* system resulted from substitutions in the amino-terminal third of HsdM and also showed enhanced modification, i.e., they were  $m^{+}$  mutants, more suggestive of defects in communication than assembly (82). Indeed, one of these mutants has been shown to make an R-M complex that is defective in restriction (D. T. F. Dryden and N. E. Murray, unpublished data).

**Subunit interactions.** Mutations conferring an  $r^{-}m^{+}$  phenotype have been identified in the *hsdS* genes of type IA and IC systems. In *EcoR124I* (189) these mutations are at the junction between a conserved region and a TRD, while in *EcoKI* they can be close to the junction (197) or at various positions within the TRD (131). A cautious interpretation of  $r^{-}$  phenotypes for *EcoKI*, however, is required by the recent discovery that mutations in *hsdM* that result in inadequate modification induce host-mediated alleviation of restriction (111) (see section on modulation of restriction activity). A mutation in *hsdS* may prevent binding of HsdS to HsdM or, as suggested by Weiserova and Firman (189), the TRD may influence the precise positioning of HsdR, perhaps in response to the methylation state of the target sequence. HsdM interacts and communicates with HsdR, but for HsdS the evidence is still unclear.

**Functional analyses of motifs.** The *hsdR* gene for *EcoKI* is now well charted, with mutations in regions that encode sequences common to all type I R-M systems (Fig. 5). These mutations indicate that each conserved region is important for restriction. Amino acid substitutions in the DEAD-box motifs do not prevent the conformational change associated with tight binding to the target sequences in the presence of AdoMet and ATP. These proteins, however, are all deficient in ATPase activity and DNA translocation, consistent with a role for the DEAD-box motifs in the coupling of ATP hydrolysis to DNA translocation (35, 37). The failure to translocate DNA was

demonstrated by an in vivo assay in which wild-type *EcoKI* translocates the T7 genome from the phage particle into the bacterial cell (57).

Conservative substitutions within the amino acid sequence characteristic of endonucleases do not block either the ATPase or DNA translocase activities of *EcoAI* (78) and *EcoKI* (37). These mutations are believed to be within a separate domain from those of the DEAD-box motifs (37), and, as expected from their sequence identity with the active sites of type II restriction endonucleases, they block the nicking and cutting activity of the R-M complex. This implies a common mechanism for the hydrolysis of phosphodiester bonds by type I and type II systems. Early reports that the ends of the DNA fragments generated by type I endonucleases are refractory to terminal labeling by the transfer of phosphate groups to the 5' ends (53) may reflect the technical problem imposed by the absence of nucleotide specificity at the site of cleavage.

**DNA translocation.** The type I R-M enzymes cleave DNA at variable positions remote from their recognition sequences (68). Electron microscopy has been used to identify possible intermediates in the reaction leading to cleavage of both linear and covalently closed plasmid DNA (52, 195). The ATP-dependent formation of DNA loops, both twisted and untwisted, has been detected. A model was proposed in which the enzyme binds to its recognition sequence, makes a second nonspecific contact with DNA, and subsequently moves the DNA past the bound complex, generating loops in an ATP-dependent process (195). It was also suggested that if type I R-M enzymes were topoisomerases, they would overcome the conformational problems encountered should the enzyme, while anchored to its target sites, maintain contact with the major groove as the DNA is pulled towards the complex. Elegant experiments using catenated plasmid DNA (171) provided direct confirmation that communication between recognition and cleavage sites stems from a process in which the enzyme follows the contour of the DNA substrate. Meanwhile, using linear DNA, Studier and Bandyopadhyay (165) provided evidence for a model in which DNA is pulled past the bound protein on both sides of the recognition sequences, and endonuclease activity results when two translocating complexes meet (Fig. 6). This model was derived from the analysis of products obtained from the digestion of T7 DNA with *EcoKI*. When the restriction reaction was synchronized by the addition of ATP to protein-DNA complexes in the presence of AdoMet, diffuse bands of DNA were detected on gels, consistent with the production of relatively discrete DNA fragments. The ends of the fragments, according to the model, would focus around the midpoints between the adjacent recognition sequences. Earlier, though less discriminating, in vivo experiments suggested the cutting of DNA between recognition sites (23). According to the Studier model, a common stimulus for cleavage of the DNA could be the hindrance of translocation (Fig. 6). Recent experiments demonstrate the stimulation of DNA cutting when two translocating enzymes from different families collide and even when a translocating enzyme encounters a Holliday junction that is unable to migrate (77).

Recently, atomic force microscopy has been used to observe the DNA translocation and cleavage process by *EcoKI* (51). In these experiments, *EcoKI* bound to two DNA target sites was seen to dimerize prior to the addition of ATP and the initiation of translocation. DNA-dependent dimerization cannot be an absolute requirement for DNA translocation, since a single target is sufficient for DNA translocation (57, 77). Dimerization of bound enzymes, however, could facilitate cooperation between two complexes, thereby enhancing restriction.

The translocation activity of *EcoKI* assayed by transfer of

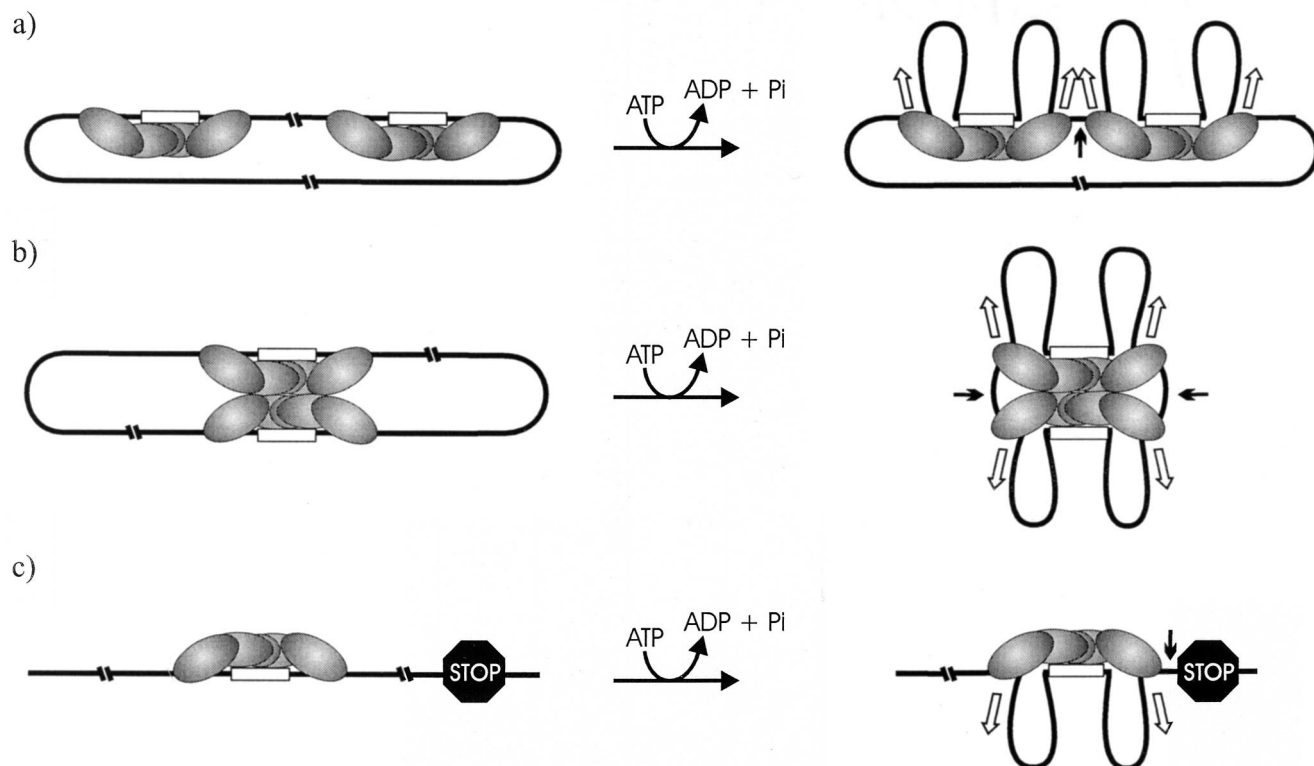


FIG. 6. ATP-dependent DNA translocation. (a) Model of Studier and Bandyopadhyay (1965). *EcoKI* bound to adjacent target sequences translocates DNA towards itself. Collision blocks translocation and stimulates the nicking of each DNA strand. Two domains of HsdR flanking the DNA are indicated. (b) In this variant, the *EcoKI* complexes have dimerized prior to translocation (51). (c) When translocation is impeded by some other protein or structure, endonuclease activity is stimulated (77).

the T7 genome from the phage particle into the bacterial cell was estimated as  $\sim 100$  bp of DNA per s (37, 57). The DEAD-box motifs are required for DNA translocation and appear to be organized within domain structures like those deduced for DNA helicases (36). These findings tally with a translocation mechanism that is dependent on helicase activity. If the enzyme follows the helical contour of the DNA while remaining bound to its target sequence, the DNA ahead of the translocating complex will be overwound (+ve supercoils) and that behind will be underwound (–ve supercoils) (36, 171). Recent experiments support this expectation and demonstrate that negative supercoils are generated by *EcoAI* in the presence of ATP and *E. coli* topoisomerase I (78a). These topological changes would impede translocation on covalently closed circular DNA in the absence of either a nicking or a topoisomerase activity. DNA nicks dependent upon the endonuclease motif of HsdR have been shown to be irrelevant to translocation in vivo (37) and in vitro (78a). Although no conventional topoisomerase activity has been observed (78a), relief by topoisomerase activity has not been ruled out, and conserved tyrosine residues have been identified within HsdR (37). Should DNA nicking and rejoining not be associated with translocation, it may be necessary to resort to a model in which the HsdR subunits are free to rotate around or detach from the methylase core of the enzyme, which remains bound to the target sequence (37, 78a).

Currently there is no direct evidence to indicate that type I R-M systems are helicases, and preliminary attempts to demonstrate helicase activity by conventional strand displacement

assays have failed (78a; G. P. Davies, personal communication). Earlier experiments, cited in reference 52 but done in 1973, approached this question by using psoralen to introduce cross-links between pyrimidine residues in the strands of T7 DNA. Low levels of psoralen blocked the activity of the RecBCD nuclease but had no detectable effect on *EcoBI*. These observations were taken to support a translocation mechanism in which *EcoBI* utilizes only the exterior of the helix, rather than strand separation (52). The effect of psoralen has not been reinvestigated using the refined techniques and substrates currently available; cross-links might trigger cutting if they impede the translocating complex.

The footprints obtained with *EcoKI* in the presence of AdoMet shorten following the addition of ATP and become similar in length to those found with  $M \cdot EcoKI$ . This could be taken as evidence for the loss of HsdR subunits, but Powell et al. (137) have shown that HsdR remains in the complexes formed between *EcoKI* and the oligonucleotide substrate of 45 bp. These authors suggested that *EcoKI* has three DNA-binding regions: a “core” region, which recognizes one target sequence, and a region on each HsdR subunit. The HsdR subunits would make tight contact with flanking DNA in the absence of cofactors, but this contact would be weakened in the presence of cofactors to allow the conformational change required for DNA translocation. Each complex would have an HsdR subunit at either side of the symmetrically arranged core of  $M \cdot EcoKI$  (Fig. 6a), and these flanking HsdR subunits would pull the DNA towards the complex from either side of the enzyme, meeting the requirement for DNA translocation in both directions.



### Assembly and Its Implications

The assembly of *EcoKI* has been analyzed in vitro (45). An assembly pathway relevant to the bacterial cell was proposed on the basis of experiments quantifying the interactions between intermediate complexes and subunits. The methyltransferase ( $M_2S_1$ ) is formed from the reversible association of  $M$  with  $M_1S_1$ . HsdR binds very tightly to both  $M_1S_1$  and  $M_2S_1$ , but the only complex with endonuclease activity is  $R_2M_2S_1$ .

The relative strengths of protein-protein interactions determined in vitro can be used, at least in part, to explain regulation of the R-M activities in vivo. It is obvious that regulation is essential when *hsd* genes are transferred to a modification-deficient recipient. Experiments fail to find evidence for transcriptional regulation of *hsdR* (106, 139). The assembly pathway could, in the following way, form the basis of posttranslational control after expression of the *hsd* genes. Initially, HsdM and HsdS will exist mainly as HsdM and  $M_1S_1$  with very little  $M_2S_1$  until the concentrations of the subunits reach a critical level. HsdR will bind  $M_1S_1$  and consequently will be unavailable for binding to  $M_2S_1$ , thereby imposing a further delay in the production of  $R_2M_2S_1$ . Recent experiments for *EcoKI* indicate that there is a lag of many generations ( $\sim 11$ ) before cells become fully modification proficient, as assessed by their ability to modify infecting  $\lambda$ 0 (S. Makovets and N. E. Murray, unpublished data) and longer,  $\sim 15$  generations, before restriction proficiency is established (138). It was proposed that any of the intermediate complexes ( $M_1S_1$  and  $R_1M_1S_1$  or HsdR) could be targets for cellular proteases (45), thereby delaying the appearance of functional restriction enzyme. However, recent studies on the host-controlled modulation of the restriction activities (see next section) question a major role for assembly pathways in controlling the restriction activities of either *EcoKI* or *EcoAI*.

All the known type I R-M enzymes are likely to assemble by a similar pathway, although the affinities of analogous protein-protein interactions may differ. For members of the type IB and IC families, the  $M_1S_1$  intermediate appears to be less prevalent than for the type IA family, but for *EcoAI* (type IB),  $R_2M_2S_1$  readily dissociates to yield the methylase and free HsdR, while for type IC enzymes, the relative affinities of HsdR for  $M_2S_1$  and  $R_1M_2S_1$  are quite different. The preferential stability of the  $R_1M_2S_1$  intermediate for *EcoR124I* (76) is currently the only explanation for the easy establishment of type IC *hsd* genes in a new host.

## BIOLOGY OF TYPE I R-M SYSTEMS

### Host-Controlled Modulation of Restriction Activity

**Self-protection.** Bacteria are assumed to tolerate the presence of a classical restriction endonuclease because the cognate modification enzyme maintains the methylation of target sequences within chromosomal DNA. A maintenance methylase does not, however, explain the long-established fact that genes encoding R-M systems are readily transferred to recipient cells in which the chromosomal DNA is unmodified. Under these circumstances, a delay in the appearance of restriction activity in the recipient cell is necessary to allow time for methylation of the unmodified chromosome (138). Other experiments demonstrate that the maintenance methyltransferase activity of an  $r^+m^+$  cell is sometimes unable to cope with the protection of unmethylated targets that arise in response to DNA damage (111). In both these cases, it is now known that there is an additional level of control over the endonuclease activity of some type I R-M systems which en-

ables the bacteria to survive in the absence of complete modification of chromosomal target sequences. This section of the review will trace the development of our current understanding of the mechanisms by which restriction activity is modulated.

Host DNA would be protected against the endonucleolytic activity of a newly acquired restriction system if the functional modification enzyme is produced before the restriction enzyme. Representatives of all three types of classical R-M systems have been shown to be equipped with promoters that could permit transcriptional regulation of the two activities. Transcriptional regulation of some of the genes encoding type II systems has been demonstrated. Genes encoding repressor-like proteins, referred to as C-proteins for control, have been identified in some instances (72, 172, 173). The C-protein for the *BamHI* system has been shown to activate efficient expression of the restriction gene and modulate the expression of the modification gene (73). Consequently, when the R-M genes are transferred to a new environment in which there is no C-protein, there will be preferential expression of the modification gene, and only after production of the C-protein will transcription of the restriction gene be activated. For type I R-M systems, despite the presence of two promoters, there is no evidence for transcriptional regulation of gene expression (95, 106, 139).

**Proteolytic control.** The heterooligomeric nature of type I R-M systems provides opportunity for the regulation of the R-M activities purely on the basis of the affinities with which different subunits bind to intermediates in the assembly pathway (45). Nevertheless, efficient transmission of the genes encoding *EcoKI* was shown to depend on a host function (139). A number of energy-dependent proteases are now known to play important regulatory roles in bacteria, and these were obvious candidates for this host function. The energy-dependent proteases identified in bacteria are often large oligomeric assemblies within which substrate recognition is imposed by one component and protease activity is imposed by another. The unfolded protein substrate is translocated to a chamber within the oligomeric complex and is then degraded processively (for reviews, see references 60 and 61). The protease specified by *clpX* and *clpP* was implicated in the regulation of restriction activity; in the absence of either ClpX or ClpP, acquisition of the *hsd* genes for *EcoKI* or for *EcoAI* led to the death of  $m^-$  recipients (110). While ClpXP is a protease, ClpX itself can function as a substrate-specific chaperone. Loss of ClpX imposed a bigger barrier to gene transfer than loss of ClpP, suggesting a dual role for ClpX. This could imply a requirement for ClpX as a chaperone in addition to its role as a component of the ClpXP protease (110).

The delay in the appearance of restriction activity following the acquisition of R-M genes by a recipient cell lacking the cognate modification activity is only one of a number of situations in which a temporary loss of restriction proficiency is required. A temporary loss of restriction proficiency, referred to as restriction alleviation (RA), also occurs in response to treatments that damage DNA (17, 38, 50, 67, 177, 178). UV light, nalidixic acid, 2-aminopurine (2-AP), and 5-bromouracil have all been shown to induce RA. The alleviation of restriction in response to these treatments, like the temporary loss of restriction proficiency associated with the establishment of a new specificity, is dependent on ClpXP (111).

DNA damage can lead to the generation of unmodified target sequences. These may arise either as the result of DNA repair via homologous recombination or as a consequence of mutations that create new target sequences. Double-strand breaks induced by external agents or by genetic lesions that lead to stalling of DNA replication are repaired by RecA-

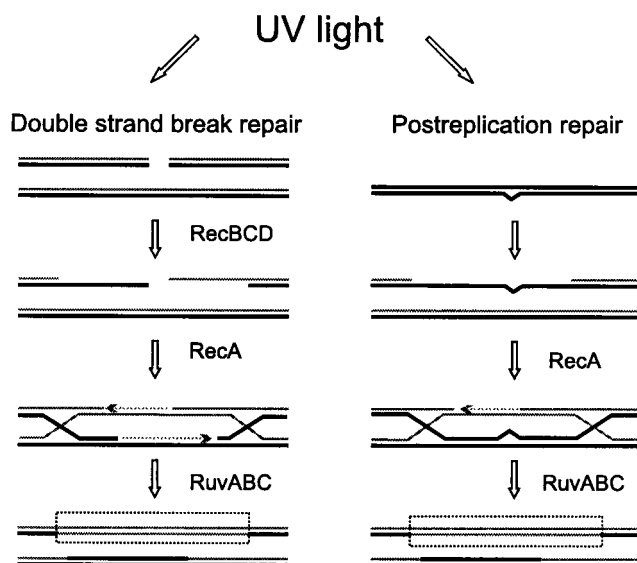


FIG. 7. Generation of unmodified target sequences following UV irradiation. Methylated strands of DNA are shown as thick lines, and unmethylated strands are shown as thin lines. Homologous recombination, involved in the repair of double-strand breaks or postreplicative repair, can generate regions of unmethylated, double-stranded DNA via annealing of two unmethylated strands (regions within boxes). In addition, the SOS mutagenesis pathway leads to new (unmodified) target sequences as the result of base changes.

dependent homologous recombination. If this recombination involves two segments of hemimethylated DNA, either the annealing of unmethylated strands or DNA synthesis may generate localized regions of unmethylated DNA (Fig. 7). In contrast, increasing the frequency of mismatches either by treatment with a mutagenic agent such as 2-AP, an analog of adenine, or by a mutator gene such as *mutD* will occasionally generate new but unmodified target sequences.

Restriction is found to be permanently (constitutively) alleviated in *dam* (49), *topA*, and *mutD* strains, and in each strain restriction is restored by a mutation in *clpX* (111). Unmodified targets in the chromosomal DNA of these mutant strains may arise continuously as the result of enhanced rates of mutation and DNA repair, and they could provide the signal for cells to protect their DNA by the alleviation of restriction activity.

The recent analysis of the role of ClpXP in RA has led to the identification of a molecular pathway that protects the bacterial chromosome against the potentially lethal effects of either an established type I R-M system or a newly acquired one. In cells encoding either *EcoKI* or *EcoAI*, RA in response to treatment with 2-AP correlates with the ClpXP-dependent loss of the HsdR subunit (111). The degradation of HsdR is prevented by either a missense mutation in *hdsR* or the absence of HsdM and HsdS, as if HsdR is degraded only when it is part of a functional complex. If the stimulus for RA is DNA breakage, then those breaks made by another system should serve as a stimulus for the degradation of the HsdR subunit of *EcoKI*. There is no evidence for degradation of the HsdR subunit of a nonfunctional *EcoKI* complex when treatment with 2-AP induces degradation of the HsdR subunit of functional *EcoAI* within the same bacterium. It appears that the signal for RA lies within *EcoKI* itself. This finding and the dependence on unmodified DNA targets suggest that HsdR is only recognized by ClpXP after the R-M complex has embarked on its restriction pathway (Fig. 8). Some step in the ATP-dependent DNA translocation could expose the target in HsdR to ClpXP (111). Strong support for this hypothesis is provided by the finding

that mutations in each of the seven DEAD-box motifs that impair translocation and ATPase activity (37) prevent ClpXP-dependent degradation of HsdR (111; V. Doronina and N. E. Murray, unpublished data), while mutations in the endonuclease motif that do not affect translocation (37) leave a complex in which the HsdR subunit remains accessible to ClpXP (V. Doronina and N. E. Murray, unpublished data). The "functional restriction complex" required as a substrate for proteolysis is one that can translocate DNA, not necessarily one that can break DNA. A remarkably specific control mechanism is inferred, effective only once the relevant restriction pathway has been initiated and able to act before any damage is inflicted on unmodified chromosomal DNA (111). The model meets an earlier suggestion of Heitman (65) that extensive DNA translocation may make type I R-M systems uniquely sensitive to DNA metabolic processes.

**Significance.** Modification has been assumed to be essential to protect the resident chromosome from restriction. However, a mutation in *hdsM* (F269G) which results in a modification-deficient, restriction-proficient *EcoKI* complex is not lethal. Residue 269 of HsdM is in motif IV (NPPF) of the catalytic domain of the methyltransferase (see section on the HsdM subunit). In the absence of methyltransferase activity, the chromosome of *E. coli* K-12 will contain ~600 unmodified targets, yet the cell survives even when repair of DNA breaks is blocked by a mutation in *recA*. Survival is dependent on

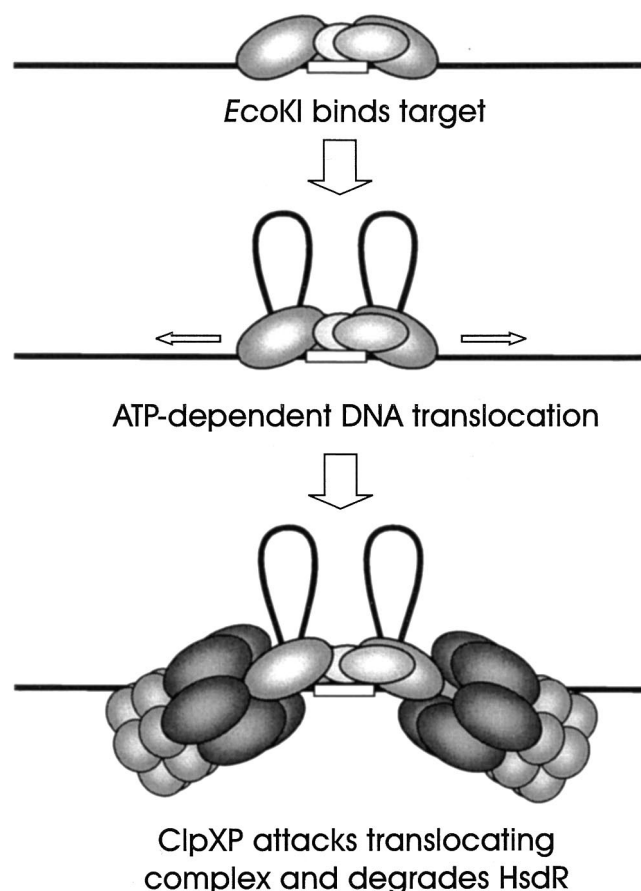


FIG. 8. Model for the mechanism of ClpXP-dependent RA (111). When a type IA or IB R-M complex binds to an unmodified target sequence, translocation is initiated and ClpXP recognizes the HsdR subunit of the complex. ClpXP degrades HsdR, preventing further translocation and endonuclease activity.

ClpXP, consistent with the ClpXP-dependent degradation of HsdR (111). RA apparently protects the bacterial chromosome from restriction in the complete absence of modification, but when analyzed by infection with unmodified  $\lambda$ , the alleviation is not complete; the EOP of  $\lambda.0$  is  $10^{-1}$ . These facts raise two new but possibly related questions. First, why does phage DNA entering the cell show some susceptibility to restriction, whereas the resident chromosome does not? Second, why do unmodified targets on the chromosome but not those of infecting  $\lambda$  DNA induce the alleviation of restriction? The classical explanation in which modification provides the basis for *EcoKI* to distinguish host DNA from foreign DNA is no longer adequate, but it can be argued that DNA modification is necessary to allow high levels of restriction activity. Important questions concerning the distinction between unmodified phage and unmodified host DNA remain to be answered. It may be relevant that the two substrates differ in their location, in their association with other proteins, or in their topology. Any of these differences could be critical for the regulation of restriction activity.

The role of ClpXP in RA has been demonstrated for type IA and IB R-M systems. Members of the type IC and ID families are also susceptible to RA in response to 2-AP (S. Makovets and N. E. Murray, unpublished data). For the plasmid-borne type IC genes, however, neither transmission by conjugation (95) nor induction of RA by 2-AP is dependent on ClpXP (S. Makovets and N. E. Murray, unpublished data). As already discussed (see Assembly and Its Implications), the assembly pathway of the  $R_2M_2S_1$  complex of type IC systems may provide a lag in the production of the endonuclease after plasmid transfer (75), but this would not readily account for RA in response to DNA damage. It is not known whether RA can be achieved by proteases other than ClpXP or by mechanisms other than proteolysis. RA has been demonstrated for a type III system and for a methylation-dependent system but is not a characteristic of type II systems (81).

The phenomenon of ClpXP-dependent RA adds a new facet to the interpretation of restriction-deficient phenotypes associated with mutations in *hsdM* or *hsdS*. Mutations in either *hsdM* (111) or *hsdS* (M. O'Neill and N. E. Murray, unpublished data) can confer an  $r^-$  phenotype as a secondary consequence of their effect on modification activity; undermethylation of chromosomal target sequences induces RA and an  $r^-$  phenotype. ClpXP therefore permits the isolation of mutations that would otherwise be lethal, but RA can obscure the true nature of the primary defect.

#### Mechanisms by Which Plasmids and Phages Avoid Restriction

Many phages and transmissible plasmids have some means of moderating their susceptibility to R-M systems. These include unusual modification of DNA, a low frequency of target sequences, and the production of a protein that interferes with one or more of the activities of an R-M system (for reviews, see references 18 and 92). DNA that enters a bacterium in a single-stranded form, as is the case for transmissible plasmids and some phages, does not avoid restriction but becomes susceptible after the synthesis of the second strand. This is true for bacterial DNA transferred by conjugation (5). Unmodified targets in newly synthesized DNA are susceptible to restriction. It is not known whether recombination is a prerequisite for unmodified targets to acquire the status of resident DNA and induce RA.

Glucosylation of the HMC residues of the DNA of T-even phages provides effective protection against most restriction

enzymes; importantly, it also identifies phage DNA so that phage-encoded enzymes can selectively degrade the host chromosome (98, 146). The *Mom* function of phage Mu converts adenine residues within a specific nucleotide sequence to  $N_6$ -(1-acetamido)adenine, a modification that is known to interfere with restriction by *EcoKI* and *EcoBI* (92).

The nucleotide sequences of some plasmids and the genomes of many phages have been analyzed for the frequencies of the target sequences recognized by those restriction enzymes that the transmissible agents might encounter (155, 190; see reference 18 for a review). Most of the statistically significant examples of underrepresentation are for the targets of type II rather than type I restriction enzymes. The apparent lack of effect on targets for type I systems may in itself be significant, since *E. coli*, the species for which most data are available, maintains a great variety of type I R-M systems encoded by allelic genes that confer different specificities (see section on Detection, Distribution, and Diversity). The promiscuous plasmid RP4 was found to be deficient in target sequences for potentially relevant type II endonucleases but not for the known type I R-M systems of *E. coli* (190). It was noted, however, that there is little evidence for the presence of promiscuous plasmids in *E. coli* strains collected prior to 1954.

Most commonly studied coliphages (18), many plasmids (14, 15, 16, 30, 148), and at least one conjugative transposon (154) specify proteins that interfere with restriction, some of which provide general protection against all known type I systems (15). The efficacy of antirestriction functions poses the intriguing question of how a protein specified by the unmodified DNA of a transmissible agent is able to act before the genome that encodes it is attacked by the restriction enzyme. Some interesting answers emerge.

The simplest solution is encountered for bacteriophage P1. This phage defends its DNA from type I restriction enzymes by proteins, designated Dar for defense against restriction, which are coinject with any encapsidated DNA (71). This mechanism allows efficient generalized transduction between strains with different type I R-M systems as well as unrestricted infection by P1 phages. How Dar protects DNA is unclear. It does not function in *trans* and therefore may remain associated with the incoming DNA. It was suggested (71) that Dar protein bound to DNA might hinder the translocation of DNA, but this does not readily explain the requirement for different Dar proteins to combat or antagonize different families of type I enzymes.

The genomes of phages T3 and T7 enter *E. coli* according to a lengthy program in which the leading genes are transcribed prior to internalization of the remainder of the genome. The antirestriction gene of these phages (0.3) is the first gene to be transcribed, and its product protects the remainder of the T7 genome from attack by type I restriction enzymes. Experimental evidence, derived from phages manipulated to include two targets for *EcoKI* in the leading region of the T7 genome, indicates that during the first few minutes of infection, the leading region is accessible to RNA polymerase but not *EcoKI*, and it was suggested that early transcription occurs in an *EcoKI*-free compartment (121), but this compartment has not been defined.

Transmissible plasmids that encode an antirestriction function, called *ard* for "alleviation of restriction DNA," resemble T3 and T7 in that the *ard* gene is located in the leading end of the DNA, but for plasmids it is single-stranded DNA that is transferred (5' to 3') by conjugation. Convincing though indirect evidence supports an elegant model for the action of *ArdA* in which the genes in the leading end of the DNA are transcribed from special promoters recognized within secondary



structures of single-stranded DNA (4, 12). This allows transcription of *ardA* and the accumulation of antirestriction protein before the transferred strand is converted into duplex DNA, the substrate for the restriction enzyme.

The entire  $\lambda$  genome enters the bacterial cell before any phage gene is expressed, and the DNA of phage  $\lambda$  is an immediate target for restriction (46). This is also true for the T7 genome when it enters the bacterium by courtesy of a  $\lambda$  capsid (57). Ral, the "restriction alleviation" function of  $\lambda$ , serves only to enhance the modification of those phages that escape restriction (105, 196). Enhanced modification is a significant advantage for a phage if the modification activity of the system, like that of *EcoKI*, is relatively ineffective on unmodified DNA. Ral affects members of the IA family but not type IB enzymes, for which modification of unmethylated DNA is efficient (105). The in vivo evidence implicates Ral as an antagonist of restriction by enhancing the modification activity of type IA systems on unmethylated DNA, but no molecular basis for this mechanism has been established. Rac, a cryptic prophage of *E. coli* K-12, specifies an analog of Ral, termed Lar; the presence of the Rac prophage enhances the EOP of  $\lambda$ .0 approximately 10-fold (105).

The 0.3 gene products of T3 and T7 are the only antirestriction functions available in significant quantities for detailed molecular analyses. These proteins, sometimes referred to as Ocr, for "overcoming restriction," bind to type I R-M enzymes and inactivate both the endonuclease and methyltransferase activities (9). Purified 0.3 protein binds tightly to *EcoKI* or *EcoBI*, preventing the enzyme from binding to DNA (9). These early experiments indicate that 1 or 2 dimers of 0.3 protein per *EcoKI* complex are sufficient to saturate the binding sites. More recent experiments indicate a value of 1 for the methyltransferase and 2 for the endonuclease (C. Atanasiu and D. T. F. Dryden, personal communication). The 0.3 protein of T3 but not that of T7 is endowed with an additional means of antagonizing type I restriction systems. This second activity degrades AdoMet, the cofactor for restriction and the methyl donor for modification (164).

Belogurov and Delver (13) noted that the 0.3 protein of T7 includes acidic sequences similar to a motif in various Ard proteins. This acidic sequence is also similar to those sequences that are common to the HsdS subunits of type I R-M systems, and these authors proposed that the antirestriction proteins compete with HsdS for assembly with the HsdM and HsdR subunits of the enzyme, but there is no evidence that the HsdS subunit of the methyltransferase is displaced by 0.3 protein in vitro (C. Atanasiu and D. T. F. Dryden, personal communication).

The glucosylated T4 genome is insensitive to type I R-M systems, but nevertheless T4 encodes a minute polypeptide, Stp, that inhibits the chromosomally encoded type IC system *EcoprrI*. This R-M system is notable because the purified complex includes an additional protein, an anticodon nuclease specified by a gene that separates *hsdR* from *hsdS*. The anticodon nuclease remains latent until Stp inactivates *EcoprrI* (see reference 80a for a detailed review).

### Detection, Distribution, and Diversity

Type I R-M systems have been detected by the barrier they impose on gene transfer, as exemplified by the original identification of the R-M systems of *E. coli* K-12 and *E. coli* B (17). Phages are a useful and sensitive test for the presence of R-M systems in laboratory strains of bacteria, but they do not provide a general screen for detecting new restriction systems. Many bacterial strains, even within the same species, and par-

ticularly when isolated from natural habitats, are unable to support the propagation of the available test phages (10).

The technical importance of type II restriction enzymes promoted intensive screens of extracts from bacterial cells for the presence of endonuclease activities that cleaved DNAs into discrete fragments. Screens for endonucleases have not been used for type I systems and are unlikely to be effective, given that the products of digestion are not discrete DNA fragments and that ATP will promote the activity of exonuclease V.

The identification of new R-M genes via sequence similarities initially relied on experimental approaches. Screening by DNA hybridization is applicable to any bacterial species, but its usefulness is limited by the availability of suitable probes. Genes encoding members of known families of type I systems may be detected by hybridization to DNA probes derived from representative *hsd* genes (10). More generally, screening databases of predicted polypeptide sequences identifies putative R-M systems in genomic sequences even when the genes would escape detection using probes derived from known families.

Biological evidence currently documents functional type I R-M systems in *Bacillus subtilis* (194), *Citrobacter freundii* (34), *Klebsiella pneumoniae* (101, 184), *L. lactis* (153), *Mycoplasma pulmonis* (47), *Pasteurella haemolytica* (66), *Salmonella enterica* (26), and *Staphylococcus aureus* (158), in addition to those originally identified in *E. coli*. All these systems have been shown to provide a barrier to phage infection, either in their original host or after cloning of their genes in *E. coli*. The three genes diagnostic of a type I system were commonly identified by coding sequences, but for the R-M system of *S. aureus* (*SauS2I*), the biological activity has not been associated with gene sequences. Screening of the present database of complete genomic sequences indicates that about half the bacterial genomes include closely linked genes that are predicted to encode the three polypeptides comprising a type I R-M enzyme (Table 1), and others include remnants, i.e., one or two *hsd* genes, commonly *hsdS*. Some strains of one species include a type I R-M system, and others do not, e.g., *B. subtilis* (182). It is clear that complex restriction enzymes are not predominantly associated with any group of bacteria: previous bias toward their presence in members of the *Enterobacteriaceae* reflects the bacteria commonly used for genetic analysis. A more surprising observation is that one strain of *Helicobacter pylori* uses more than 1% of its small genome to encode R-M systems, three of which are type I (181).

For *E. coli*, the genomes of wild-type isolates have been screened with probes specific for each of the three families of chromosomally encoded type I R-M systems (IA, IB, and ID) identified in the *Enterobacteriaceae*. Thirty-seven strains from the *E. coli* reference library (ECOR strains) were tested, and 17 included sequences identified by one of the probes. No genome was found to hybridize with more than one probe, as expected if the *hsd* genes are alleles of one locus and consistent with the subdivision of the systems into discrete families using DNA hybridization as a test for sequence similarity (11). The *serB*-linked *hsd* genes identified in eight serotypes of *S. enterica* (26) include representatives of the three families, but the *hsd* genes in three of the eight strains were not detected with any probe (10, 180). A summary of the classification of the type I R-M systems in the *Enterobacteriaceae* is given in Table 2. The allocation to families has not always been confirmed by complementation tests, but interspecific tests show that defects in *EcoKI* (IA) can be complemented by *StyLTHI* (185) and those in *StySBLI* (ID) can be complemented by subunits of both *EcoR9I* (A. J. B. Titheradge and N. E. Murray, unpublished data) and *KpnAI* (J. Ryu, personal communication).

The two strains of *H. pylori* for which complete genomic

TABLE 1. Summary of evidence for *hsd* genes in genomic sequences

Organism <sup>a</sup>	<i>hsdR</i> / <i>hsdM</i> / <i>hsdS</i> present	Orphan genes <sup>c</sup>
<i>Aeropyrum pernix</i> K1		
<i>Aquifex aeolicus</i> VF5		
<i>Archaeoglobus fulgidus</i>	+	
<i>Bacillus subtilis</i> 168	— <sup>b</sup>	
<i>Borrelia burgdorferi</i>		
<i>Chlamydia pneumoniae</i> CWL029		
<i>Chlamydia trachomatis</i> D/UW-3/CX		
<i>Escherichia coli</i> K-12 MG1655	+	
<i>Haemophilus influenzae</i> Rd	+	
<i>Helicobacter pylori</i> 26695	+	+
<i>Helicobacter pylori</i> J99	+	+
<i>Methanobacterium thermoautotrophicum</i> delta H	+	
<i>Methanococcus jannaschii</i>	+	+
<i>Mycobacterium tuberculosis</i> H37Rv		+
<i>Mycoplasma genitalium</i> G37		+
<i>Mycoplasma pneumoniae</i> M129	+	+
<i>Pyrococcus abyssi</i> Orsay	+	
<i>Pyrococcus horikoshii</i> OT3		
<i>Rickettsia prowazekii</i> Madrid E		
<i>Synechocystis</i> sp. strain PCC6803	— <sup>b</sup>	
<i>Thermotoga maritima</i>		
<i>Treponema pallidum</i>		

<sup>a</sup> Strains for which the complete genome sequence and annotation are available.

<sup>b</sup> Type I R-M genes have been identified in other strains of these species.

<sup>c</sup> Only one or two of the three genes were detected.

sequences are available, strains 26695 (181) and J99 (3), each include coding sequences for three putative type I R-M systems. The predicted polypeptide sequences for systems within a strain, either 26695 or J99, show only modest levels of amino acid identity (20 to 40%). The three systems in strain 26695 have homologs in J99, based on amino acid identities (80 to 90%) of HsdR and HsdM. Some of the HsdS sequences are sufficiently short to suggest that they are not functional. The HsdS sequences of one pair of homologs (HP0462 and jhp0414) are of normal length for known type I systems. Their sequences have high identity in the amino-terminal TRD but low identity in the carboxy-terminal TRD, consistent with allelic genes that confer a different specificity.

A modified concept of families is emerging from studies of *L. lactis*, in which it appears that HsdS subunits can be exchanged between two families of type I R-M systems. The HsdR and HsdM subunits specified by two plasmids have the high levels of amino acid identity (~90%) expected for members of one family (153). The sequences of these plasmid-encoded subunits, however, show only 40% identity with those specified by chromosomally encoded genes. These low levels of identity between the plasmid- and chromosomally encoded systems of *L. lactis* are consistent with the concept of two families of enzymes. In contrast, all the sequence data for HsdS subunits, whether plasmid or chromosomal in origin, indicate a single family of enzymes with differences in HsdS sequences found principally within the TRDs. The conserved regions of HsdS subunits include those sequences that interact with HsdM, subunits that are well conserved within a family (62, 86, 109, 189). Schouler et al. (153) note that, irrespective of the family designation, the C-terminal parts of the HsdM subunits of the lactococcal systems have a common sequence that could identify a region involved in the association of HsdM with HsdS. Complementation tests indicate that the HsdS subunits of plasmid systems can form functional complexes with the

chromosomally encoded HsdM and HsdR subunits. The present data suggest that a reservoir of independent *hsdS* genes on plasmids provides an effective "combinational" system for varying the target specificity of the catalytic subunits specified by the host.

The type I R-M systems of *M. pulmonis* illustrate the most sophisticated system for the expression of different specificities; in this case, variation is found among the descendants of one bacterium (48). The bacterial chromosome includes two *hsd* loci, each of which contains two *hsdS* genes flanking *hsdM* and *hsdR*. Site-specific DNA inversion can occur using any of three sites within the *hsdS* genes. There is only one promoter for each *hsd* locus, so that only one orientation permits transcription of a functional *hsd* operon, but the alternative sites of inversion create a family of functional HsdS subunits by gene rearrangements. This switching of specificities is reminiscent of the phase variation of virulence determinants.

### Evolution

Of all known R-M systems, those classified as type I appear to be uniquely suited for diversification of sequence specificity. In part this is because a single subunit is responsible for the sequence specificity of both enzyme activities. Importantly, therefore, the specificities of the endonuclease and methyltransferase change concomitantly. In addition, some, perhaps all, type I R-M systems have coevolved with their hosts to ensure that neither the acquisition of such a system nor a change in the specificity of an established system leads to endonucleolytic attack of the host chromosome (110, 130). While both the shared specificity subunit and the control of restriction activity facilitate change of sequence specificity, the scope for diversification is enhanced by the presence of two different TRDs within one specificity subunit. The two TRDs

TABLE 2. Type I R-M systems identified in members of the family *Enterobacteriaceae*

R-M family	Enzyme	Target sequence	Reference(s)
IA	<i>EcoKI</i> <sup>a</sup>	AAC(N <sub>6</sub> )GTGC	79
	<i>EcoBI</i> <sup>b</sup>	TGA(N <sub>8</sub> )TGCT	147
	<i>EcoDI</i>	TTA(N <sub>7</sub> )GTCY	125
	<i>StyLTHI</i> <sup>c</sup>	GAG(N <sub>6</sub> )RTAYG	126
	<i>StySPI</i>	AAC(N <sub>6</sub> )GTRC	126
	<i>EcoR51</i>	ND <sup>d</sup>	11
	<i>EcoR10I</i> <sup>e</sup>	ND	11
	<i>EcoR23I</i>	ND	11
IB	<i>EcoAI</i>	GAG(N <sub>7</sub> )GTCA	91, 168
	<i>EcoEI</i>	GAG(N <sub>7</sub> )ATGC	33
	<i>CfrAI</i>	GCA(N <sub>6</sub> )GTGG	80
	<i>StySKI</i>	CGAT(N <sub>7</sub> )GTTA	179
	<i>StySTI</i>	ND	10
	<i>EcoR17I</i>	ND	11
IC	<i>EcoR124I</i>	GAA(N <sub>6</sub> )RTCG	142
	<i>EcoDXXI</i>	TCA(N <sub>7</sub> )RTTC	64
	<i>EcoprI</i>	CCA(N <sub>7</sub> )RTGC	183
ID	<i>StySBLI</i>	CGA(N <sub>6</sub> )TACC	180; Titheradge and Murray, unpublished
	<i>EcoR9I</i>	ND	11
	<i>KpnAI</i>	ND	101

<sup>a</sup> Strains ECOR12 and -24 have type IA systems with the same specificity.

<sup>b</sup> Strains ECOR14, 42 and 70 have type IA systems with the same specificity.

<sup>c</sup> Referred to as *StySBI* in the early literature.

<sup>d</sup> ND, not done.

<sup>e</sup> Strain ECOR25 has a type IA system with the same specificity.

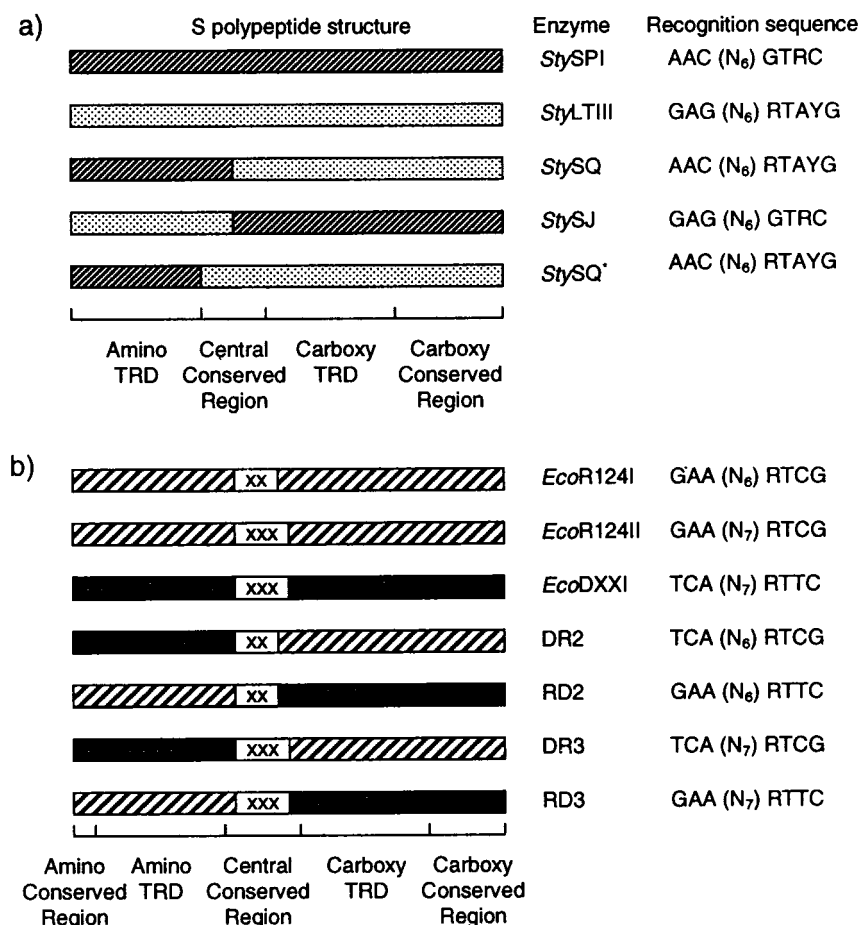


FIG. 9. Evolution of type I R-M systems with new specificities. (a) Recombination between *hsdS* genes produces hybrid genes and chimeric S polypeptides. *Sty*SPI and *Sty*LIII are naturally occurring type I R-M systems (see Table 2). *Sty*SQ and *Sty*SJ have hybrid *hsdS* genes (55, 56). The regions originating from *Sty*SPI are hatched, and those originating from *Sty*LIII are stippled. Reassortment of the target recognition domains (TRDs) accordingly gave rise to recombinant recognition sequences (56, 127). Site-directed mutagenesis of the central conserved region of the *Sty*SQ *hsdS* gene produced *Sty*SQ\*, comprising only the amino-terminal variable region from *Sty*SPI and the remainder from *Sty*LIII. The *Sty*SQ\* target sequence confirms that the amino-terminal variable region is in fact a TRD responsible for recognition of the trinucleotide component of the sequence (33). (b) Sequence specificity may also be altered by changing the length of the nonspecific spacer of the target sequence. The S polypeptides of *Eco*R124I and *Eco*R124II differ only in the number of times a short amino acid motif (X = TAEI) is repeated within their central conserved regions (142), resulting in extension of the spacer in the target sequence from six nucleotides (N<sub>6</sub>) in *Eco*R124I to N<sub>7</sub> in *Eco*R124II. The recognition sequence of *Eco*DXXI also contains a nonspecific spacer of seven nucleotides, corresponding to three TAEI repeats in its S polypeptide (64). Chimeric S polypeptides recognize the predicted target sequences (64). Modified from a figure in Barcus and Murray (10) with permission.

of a homodimeric type II restriction endonuclease dictate a rotationally symmetrical target sequence. In contrast, for known type I systems, the two TRDs recognize different sequences, and it was shown that specificity can be changed not only by reassortment of TRDs, but also by altering the spacing between them (Fig. 9). Chance examples of these two routes to diversification were encountered in the laboratory. The new R-M system found as the by-product of P1-mediated transduction (25) arose from a genetic exchange between different *hsdS* genes; a cross-over within the conserved DNA sequences in the interval separating the variable regions that specify TRDs generated a new combination of TRDs (55, 127). The second serendipitous change in DNA specificity arose in a type IC system (70). In this family, a short duplication of 12 bp within the conserved sequence of an *hsdS* gene provides the opportunity for mispairing and unequal crossing-over. The target sequences of the new specificity are separated by an extra base pair as the consequence of a triplication of the 12-bp repeat (142).

New specificities as the result of the reassortment of TRDs depend upon a natural reservoir of related *hsdS* genes avail-

able for genetic recombination. The target sequences of five members of the IA family are known, and only two have a common component (Table 2). However, the generation of a TRD with a different specificity has never been witnessed, despite experiments designed to select for a relaxation in the specificity of *Eco*KI [5' AAC(N<sub>6</sub>)GTGC] to *Sty*SPI [5' AAC(N<sub>6</sub>)GTRC, where R is a purine] (N. E. Murray, unpublished data). It seems probable that the evolution of a new specificity will require an appropriate combination of amino acid changes.

The allelic *hsdS* genes of a family of type I systems retain a record of some of the evolutionary changes associated with the diversification of sequence specificity, while the HsdM and HsdR polypeptides are conserved to associate with variant HsdS subunits. Nevertheless, detailed comparisons of nucleotide sequences of the *hsdM* genes of *E. coli* K-12 and B revealed that they are far more divergent in terms of both synonymous and nonsynonymous changes than any of 29 other pairs of genes available for comparison (156). It is readily argued that the diversity in specificity reflects the action of



**N-terminus**

Central conserved region (ccr)

```

KAPELRFDGFTDDWEERKLLD - N-TRD - EQRKIGLFFKQLDDTIVLHQRKLDLLKEQKKGYLQKMF PKNGS
* ***** * ***** ** ** ***** ** ***** ***** *****
KIPELRFAEFADDWEERKLGE - C-TRD - EOKOIGSFFKOLDNTITLHORKLDLLKEOKKGFLOKMF

```

ccr continued

**C-terminus**

FIG. 10. Repeated amino acid sequences in an HsdS polypeptide of *L. lactis*. The sequence (153) has been aligned to emphasize the repeats flanking the predicted target recognition domains.

natural selection, and it is tempting to speculate that the high level of divergence between the *hsdM* alleles also reflects selection. Levin (102) has pointed out that phage-mediated selection of bacterial R-M systems is likely to be frequency dependent, thereby favoring the retention of rare genotypes and the promotion of diversity. Phage-mediated selection is expected to act directly on *hsdS*. Some new *hsdS* alleles will be acquired by recombination, and therefore the high divergence at *hsdM* may result from its tight linkage with *hsdS*. Allelic diversity at the *hsd* locus of *E. coli* parallels the polymorphisms noted for cell surface antigens where frequency-dependent selection has been invoked for favoring rare antigens (21).

The high level of sequence divergence within a family (123, 156), the abnormal G+C content within *hsdS* genes (39), and the alternative families of *hsd* genes encoded by allelic genes (11) support the concept of lateral transfer of *hsd* genes. Lateral transfer of genes of very dissimilar nucleotide sequence does not explain their common chromosomal location or, in the case of the ID family, a different gene order (180). Would one expect the maintenance of sufficient DNA homology in flanking regions to direct insertion of different families of *hsd* genes at the same site? At one locus in *B. subtilis*, strain-specific variation identifies the alternative to genes for a type I R-M system as those specifying type II systems (182). In this case the transfer of alternative sets of genes to *B. subtilis* 168 has been shown to rely on flanking homology (T. A. Trautner, personal communication), but why would closely related bacteria have such different gene sequences at a common location?

The hyphenated, symmetrical target sequences detected for mutant type I systems in which the *hsdS* genes are truncated (Fig. 3a) imply that two "half" specificity subunits (HsdS\*) within one TRD can substitute for a single wild-type HsdS with two TRDs (1, 113). This is consistent with a complex of  $R_2M_2S^*_{2_2}$  rather than the normal  $R_2M_2S_1$ . Sequence analyses provide evidence for the origin of the present *hsdS* gene by gene duplication. The evidence for the type IA family is relatively weak (8), probably because the first part of the gene has suffered a deletion, but the evidence remains strong for members of the IB family (80). In this case a sequence of 192 nucleotides at the beginning of *hsdS* is repeated in the center, with 55 of the 65 encoded amino acids being identical. The recently published sequences of the *hsdS* genes in the gram-positive bacterium *L. lactis* (153) reinforce the evidence for gene duplication, identifying a split repeat within which 47 of 59 encoded amino acids are identical (Fig. 10).

Ancestral enzymes, in which HsdS has only a single TRD, probably had the composition  $R_2M_2S_2$ . Some unusual R-M systems are now known to comprise an endonuclease that is like a type II enzyme but is accompanied by a modification enzyme more akin to a type I methyltransferase. The *AhdI*

methyltransferase, for example, is made up of subunits resembling HsdM and a half HsdS. The enzyme recognizes a hyphenated, symmetrical target sequence, and the subunits have more sequence similarity with type I systems than has been found for classical type II methyltransferases (G. G. Wilson, personal communication).

At the level of DNA sequence, little evidence exists for homology between either the *hsd* alleles of different families of enzymes or even the variable regions of *hsdS* genes within a family unless the variable regions specify the same target sequence. Nevertheless, evidence for homology remains in the nucleotide sequence of variable regions of *hsdS* genes from different families when the TRDs specify the same target sequence (32).

It seems significant that the amino acid sequences and predicted secondary structures of type I enzymes have led to model structures similar to those deduced experimentally for type II enzymes, in particular, the active site of type I methyltransferases (44) and the regions of the TRDs located close to the DNA target (131, 166). Irrespective of family, or even type, the evidence from the protein domains is consistent with the concept that R-M systems have common building blocks.

### Relevance to Bacteria

R-M systems are both widespread and diverse in nature, yet there is no indication that a strain of *E. coli* deprived of its restriction system is disadvantaged under normal laboratory conditions. It is evident from surveys of *E. coli* that allelic variation provides diversity (11), and this diversity is therefore expected to have relevance within the species. The classical explanation for the presence of R-M systems is that restriction is a barrier to phage infection. There is much evidence for the impact of R-M systems on phages (see section on Mechanisms by Which Plasmids and Phages Avoid Restriction), but there is no proof that bacteria maintain R-M enzymes to protect themselves against phages. Experiments indicate that bacteria sharing a habitat with phages rapidly acquire resistance to infection, and restriction-proficient bacteria encoding a system with a novel specificity are likely to be at an advantage only when colonizing a new habitat (89, 90, 102). This advantage may be sufficient to impose frequency-dependent selection for diversity.

As early as 1973, it was realized that the DNA breaks induced by restriction endonucleases could be recombinogenic (S. Lederberg, cited in reference 143), and it has been pointed out that because the breaks made by type I enzymes are not within the target sequences, any gene has a chance to be close to a break (52a, 140).

For a time there appeared to be a conflict between a recombinogenic role for RecBCD, alias exonuclease V, in salvaging

the products of restriction by recombination and its proven importance in the degradation of the linear DNA fragments produced following the restriction of phage (157). This apparent contradiction was resolved by an appreciation of the "recombinator" sequence Chi, an 8-base sequence missing in phage  $\lambda$  and well represented in the bacterial chromosome (140, 163). RecBCD travels along the DNA from an end, and on encountering an appropriately oriented Chi sequence, DNA degradation is attenuated and recombination activity is stimulated (40, 162). Following conjugation, most recombination is believed to result in the incorporation of long "chunks" of donor DNA (159), with the first exchange occurring within the DNA close to the origin of transfer. This recombination is mediated by the RecBCD pathway. Fragmentation of DNA by restriction should increase the opportunity for the incorporation of short, rather than long, stretches of DNA. Sequence comparisons of chromosomal DNA from several *E. coli* reference strains show that members of these natural isolates differ by short lengths of nucleotide sequence (119). Analyses of gene transfer between natural isolates of *E. coli* and *E. coli* K-12 by transduction and conjugation document the influence of R-M systems on the pattern of DNA fragments incorporated into the recipient chromosome, creating mosaic genomes of short sequences originating from different strains (112, 120). R-M systems therefore will influence the flux of genetic material between bacterial populations, enhancing the opportunity for the acquisition of advantageous coding sequences in the absence of deleterious ones (120).

The groups of Kobayashi (87, 128) and Ehrlich (94) have provided experimental evidence that type II R-M genes behave as "selfish" entities; the R-M genes are not readily lost from the host cell but maintain the presence of the plasmids that encode them. These results imply that the loss of type II R-M genes leads to descendants that die because they retain some endonuclease activity at a time when there is insufficient modification enzyme to protect all the target sequence within the chromosome. An explanation for diversity of sequence specificity in the case of selfish R-M genes stems from the finding that one system excludes another if both recognize the same target sequence (96). In other words, systems with the same target sequence behave like plasmids in the same compatibility group, and selection could arise from competition for recognition sequences rather than being driven by phages.

The type I systems studied, *EcoKI*, *EcoR124I*, and *EcoAI*, do not behave like type II systems in many relevant respects. Loss of the genes specifying a type I system has no detectable effect on bacterial viability (95, 110, 130). For *EcoKI*, a mutation that blocks modification activity but leaves *hsdR* functional, e.g., deletion of *hsdM* and *hsdS* or a missense mutation in *hsdS*, can be detrimental to viability but only in the absence of both ClpXP and some other unknown function, and even then only in stationary-phase cultures (S. Makovets and N. E. Murray, unpublished data). Apparently, in the absence of ClpXP, a mutation in *hsdS* can lead to progeny in which dilution of the modification enzyme, following cell growth and division, results in the inadequate modification of chromosomal DNA, thus generating targets susceptible to residual restriction endonuclease. Normally, ClpXP-dependent RA successfully protects the bacterial chromosome from damage by the endonuclease activity of *EcoKI*. On the current model (Fig. 8), any *EcoKI* molecule caught in the act of translocating chromosomal DNA is prevented from breaking the DNA. For *EcoKI* and *EcoAI*, a refined target-specific system exercises tight control of endonuclease activity. This is probably particularly necessary for a system like *EcoKI*, in which modification of unmethylated DNA is very inefficient.

It has been suggested that the proteolytic control of endonuclease activity permits efficient variation of the specificity of type I R-M systems (111). The organization of type I enzymes, the bipartite asymmetrical nature of the target sequence, and the host-mediated alleviation of restriction all emphasize the importance of diversity. Selection is presumed to contribute to the maintenance of allelic diversity. R-M systems in general may have evolved under a variety of pressures, but it is tantalizingly difficult to prove the biological relevance of such varied and sophisticated molecular systems.

## CONCLUSION AND FUTURE DIRECTIONS

The characterization of wild-type enzymes is now enhanced by the analysis of mutants and supplemented by comparisons with simpler proteins of known structure. Such comparisons are often suggestive of mechanisms for the enzymatic activities, but so far they have been unable to address the key question of how the subunits of the methyltransferase communicate to the HsdR subunits that the DNA is a substrate for DNA translocation rather than for modification. Mutations in *hsdR* which block endonuclease activity will facilitate the study of DNA translocation. Mutations in *hsdM* and *hsdS* that confer an  $r^- m^+$  phenotype and their allelic-specific suppressor mutations should help to identify sequences involved in the interaction of subunits and the communication between them. Currently, the interpretation of all genetic analyses remains hampered by the absence of direct structural information. A key challenge, therefore, is to obtain crystals suitable for X-ray diffraction.

The foremost biological question raised by the recent demonstration of the importance of RA is that of how *E. coli* distinguishes unmodified targets in resident chromosomal DNA from those presented by incoming DNA; unmodified targets in the chromosome elicit RA, but those in an infecting phage generally evoke restriction. The conditions that provoke the degradation of HsdR in vitro have not been reported. An understanding of what is necessary to make HsdR susceptible to ClpXP may provide the clue to the distinction between resident and incoming DNA.

## ACKNOWLEDGMENTS

I express my appreciation to those who have provided data and discussion, most particularly past and present members of my laboratory. I thank Tom Bickle, Laurie Cooper, Graham Davies, Victoria Doronina, David Dryden, Alexander Gann, Svetlana Makovets, Kenneth Murray, Mary O'Neill, Lynn Powell, Annette Titheradge, and Brian Wulks for constructive comments on the manuscript; Graham Davies and Svetlana Makovets for graphics; and Natalie Honeyman for her patient and careful preparation of the manuscript. I am indebted to Annette Titheradge for her analyses of genomic sequences and to REBASE for facilitating access to the data. Finally, I wish to acknowledge the Medical Research Council of Great Britain, who have supported my research since 1967, and Naomi Franklin, who aroused my interest in restriction systems in 1962 when she described the fundamental experiments of Daisy Dussoix and Werner Arber.

## REFERENCES

1. Abadjieva, A., J. Patel, M. Webb, V. Zinkevich, and K. Firman. 1993. A deletion mutant of the type IC restriction endonuclease *EcoR124I* expressing a novel DNA specificity. *Nucleic Acids Res.* **21**:4435-4443.
2. Allen, D. J., A. Makhov, M. Grilley, J. Taylor, R. Thresher, P. Modrich, and J. D. Griffith. 1997. MutS mediates heteroduplex loop formation by a translocation mechanism. *EMBO J.* **16**:4467-4476.
3. Alm, R. A., L. S. Ling, D. T. Moir, B. L. King, E. D. Brown, P. C. Doig, D. R. Smith, B. Noonan, B. C. Guild, B. L. deJonge, G. Carmel, P. J. Tummino, A. Caruso, M. Uria-Nickelsen, D. M. Mills, C. Ives, R. Gibson, D. Merberg, S. D. Mills, Q. Jiang, D. E. Taylor, G. F. Vovis, and T. J. Trust. 1999. Genomic-sequence comparison of two unrelated isolates of the human

- gastric pathogen *Helicobacter pylori*. *Nature* **397**:176–180.
4. Althorpe, N. J., P. M. Chilley, A. T. Thomas, W. J. Brammar, and B. M. Wilkins. 1999. Transient transcriptional activation of the IncI plasmid anti-restriction gene (*ardA*) and SOS inhibition gene (*psiB*) early in conjugating recipient bacteria. *Mol. Microbiol.* **31**:133–142.
  5. Arber, W. 1965. Host-controlled modification of bacteriophage. *Annu. Rev. Microbiol.* **19**:365–378.
  6. Arber, W., and D. Dussoix. 1962. Host specificity of DNA produced by *Escherichia coli*. I. Host controlled modification of bacteriophage lambda. *J. Mol. Biol.* **5**:18–36.
  7. Arber, W., and D. Wauters-Willems. 1970. Host specificity of DNA produced by *Escherichia coli*. XII. The two restriction and modification systems of strain 15T-. *Mol. Gen. Genet.* **108**:203–217.
  8. Argos, P. 1985. Evidence for a repeating domain in type I restriction enzymes. *EMBO J.* **4**:1351–1355.
  9. Bandyopadhyay, P. K., F. W. Studier, D. L. Hamilton, and R. Yuan. 1985. Inhibition of the type I restriction-modification enzymes *EcoB* and *EcoK* by the gene 0.3 protein of bacteriophage T7. *J. Mol. Biol.* **182**:567–578.
  10. Barcus, V. A., and N. E. Murray. 1995. Barriers to recombination: restriction, p. 31–58. In S. Baumberg, J. P. W. Young, S. R. Saunders, and E. M. H. Wellington (ed.), *Population genetics of bacteria*. Cambridge University Press, Cambridge, United Kingdom.
  11. Barcus, V. A., A. J. B. Titheradge, and N. E. Murray. 1995. The diversity of alleles at the *hsd* locus in natural populations of *Escherichia coli*. *Genetics* **140**:1187–1197.
  12. Bates, S., R. A. Roscoe, N. J. Althorpe, W. J. Brammar, and B. M. Wilkins. 1999. Expression of leading region genes on IncI plasmid ColIb-P9: genetic evidence for single-stranded DNA transcription. *Microbiology* **145**:2655–2662.
  13. Belogurov, A. A., and E. P. Delver. 1995. A motif conserved among the type I restriction-modification enzymes and antirestriction proteins: a possible basis for mechanism of action of plasmid-encoded antirestriction functions. *Nucleic Acids Res.* **23**:785–787.
  14. Belogurov, A. A., E. P. Delver, and O. V. Rodzevich. 1992. IncN plasmid pKM101 and IncI plasmid ColIb-P9 encode homologous antirestriction proteins in their leading regions. *J. Bacteriol.* **174**:5079–5085.
  15. Belogurov, A. A., E. P. Delver, and O. V. Rodzevich. 1993. Plasmid pKM101 encodes two nonhomologous antirestriction proteins (ArdA and ArdB) whose expression is controlled by homologous regulatory sequences. *J. Bacteriol.* **175**:4843–4850.
  16. Belogurov, A. A., T. Yussifov, V. U. Kotova, and G. B. Zavilgelsky. 1985. The novel gene(s) ARD of plasmid pKM101: alleviation of EcoK restriction. *Mol. Gen. Genet.* **198**:509–513.
  17. Bertani, G., and J. J. Weigle. 1953. Host-controlled variation in bacterial viruses. *J. Bacteriol.* **65**:113–121.
  18. Bickle, T. A., and D. H. Kruger. 1993. Biology of DNA restriction. *Microbiol. Rev.* **57**:434–450.
  19. Bickle, T. A., C. Brack, and R. Yuan. 1978. ATP-induced conformational changes in the restriction endonuclease from *Escherichia coli* K-12. *Proc. Natl. Acad. Sci. USA* **75**:3099–3103.
  20. Bird, L. E., H. S. Subramanya, and D. B. Wigley. 1998. Helicases: a unifying structural theme? *Curr. Opin. Struct. Biol.* **8**:14–18.
  21. Biseric, M., J. Y. Feutrier, and P. R. Reeves. 1991. Nucleotide sequences of the *gnd* genes from nine natural isolates of *Escherichia coli*: evidence of intragenic recombination as a contributing factor in the evolution of the polymorphic *gnd* locus. *J. Bacteriol.* **173**:3894–3900.
  22. Boyer, H. W., and D. Roulland-Dussoix. 1969. A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. *J. Mol. Biol.* **41**:459–472.
  23. Brammar, W. J., N. E. Murray, and S. Winton. 1974. Restriction of lambda trp bacteriophages by *Escherichia coli* K. *J. Mol. Biol.* **90**:633–647.
  24. Buhler, R., and R. Yuan. 1978. Characterization of a restriction enzyme from *Escherichia coli* K carrying a mutation in the modification subunit. *J. Biol. Chem.* **253**:6756–6760.
  25. Bullas, L. R., C. Colson, and A. Van Pel. 1976. DNA restriction and modification systems in *Salmonella*. SQ, a new system derived by recombination between the SB system of *Salmonella typhimurium* and the SP system of *Salmonella potsdam*. *J. Gen. Microbiol.* **95**:166–172.
  26. Bullas, L. R., C. Colson, and B. Neufeld. 1980. Deoxyribonucleic acid restriction and modification systems in *Salmonella*: chromosomally located systems of different serotypes. *J. Bacteriol.* **141**:275–292.
  27. Burckhardt, J., J. Weisemann, and R. Yuan. 1981. Characterization of the DNA methylase activity of the restriction enzyme from *Escherichia coli* K. *J. Biol. Chem.* **256**:4024–4032.
  28. Chen, A., L. M. Powell, D. T. F. Dryden, N. E. Murray, and T. Brown. 1995. Tyrosine 27 of the specificity polypeptide of *EcoKI* can be UV crosslinked to a bromodeoxyuridine-substituted DNA target sequence. *Nucleic Acids Res.* **23**:1177–1183.
  29. Cheng, X. 1995. Structure and function of DNA methyltransferases. *Annu. Rev. Biophys. Biomol. Struct.* **24**:293–318.
  30. Chilley, P. M., and B. M. Wilkins. 1995. Distribution of the *ardA* family of antirestriction genes on conjugative plasmids. *Microbiology* **141**:2157–2164.
  31. Cooper, L. P., and D. T. F. Dryden. 1994. The domains of a type I DNA methyltransferase: interactions and role in recognition of DNA methylation. *J. Mol. Biol.* **236**:1011–1021.
  32. Cowan, G. M. 1988. A new family of type I restriction and modification systems. Ph.D. thesis. University of Edinburgh, Scotland.
  33. Cowan, G. M., A. A. Gann, and N. E. Murray. 1989. Conservation of complex DNA recognition domains between families of restriction enzymes. *Cell* **56**:103–109.
  34. Daniel, A. S., F. V. Fuller-Pace, D. M. Legge, and N. E. Murray. 1988. Distribution and diversity of *hsd* genes in *Escherichia coli* and other enteric bacteria. *J. Bacteriol.* **170**:1775–1782.
  35. Davies, G. P., L. M. Powell, J. L. Webb, L. P. Cooper, and N. E. Murray. 1998. *EcoKI* with an amino acid substitution in any one of seven DEAD-box motifs has impaired ATPase and endonuclease activities. *Nucleic Acids Res.* **26**:4828–4836.
  36. Davies, G. P., I. Martin, S. S. Sturrock, A. Cronshaw, N. E. Murray, and D. T. F. Dryden. 1999. On the structure and operation of type I DNA restriction enzymes. *J. Mol. Biol.* **290**:565–579.
  37. Davies, G. P., P. Kemp, I. J. Molineux, and N. E. Murray. 1999. The DNA translocation and ATPase activities of restriction-deficient mutants of *EcoKI*. *J. Mol. Biol.* **292**:787–796.
  38. Day, R. S. 1977. UV-induced alleviation of K-specific restriction of bacteriophage lambda. *J. Virol.* **21**:1249–1251.
  39. Dila, D., E. Sutherland, L. Moran, B. Slatko, and E. A. Raleigh. 1990. Genetic and sequence organization of the *mcrBC* locus of *Escherichia coli* K-12. *J. Bacteriol.* **172**:4888–4900.
  40. Dixon, D. A., and S. C. Kowalczykowski. 1993. The recombination hotspot X is a regulatory sequence that acts by attenuating the nuclease activity of the *E. coli* RecBCD enzyme. *Cell* **73**:87–96.
  41. Dreier, J., M. P. MacWilliams, and T. A. Bickle. 1996. DNA cleavage by the type IC restriction-modification enzyme *EcoRI*24II. *J. Mol. Biol.* **264**:722–733.
  42. Dryden, D. T. F. 1999. Bacterial DNA methyltransferases, p. 283–340. In X. Cheng and R. M. Blumenthal (ed.), *S-Adenosyl-methionine-dependent methyltransferases: structure and function*. World Scientific Publishing, Singapore.
  43. Dryden, D. T. F., L. P. Cooper, and N. E. Murray. 1993. Purification and characterization of the methyltransferase from the type I restriction and modification system of *Escherichia coli* K12. *J. Biol. Chem.* **268**:13228–13236.
  44. Dryden, D. T. F., S. S. Sturrock, and M. Winter. 1995. Structural modelling of a type I DNA methyltransferase. *Nat. Struct. Biol.* **2**:632–635.
  45. Dryden, D. T. F., L. P. Cooper, P. H. Thorpe, and O. Byron. 1997. The *in vitro* assembly of the *EcoKI* type I DNA restriction/modification enzyme and its *in vivo* implications. *Biochemistry* **36**:1065–1076.
  46. Dussoix, D., and W. Arber. 1962. Host specificity of DNA produced by *Escherichia coli*. II. Control over acceptance of DNA from infecting phage lambda. *J. Mol. Biol.* **5**:37–49.
  47. Dybvig, K., and H. Yu. 1984. Regulation of a restriction-modification system via DNA inversion in *Mycoplasmata pulmonis*. *Mol. Microbiol.* **12**:547–560.
  48. Dybvig, K., R. Sitaraman, and C. T. French. 1998. A family of phase-variable restriction enzymes with differing specificities generated by high-frequency gene rearrangements. *Proc. Natl. Acad. Sci. USA* **95**:13923–13928.
  49. Efimova, E. P., E. P. Delver, and A. A. Belogurov. 1988. Alleviation of type I restriction in adenine methylase (*dam*) mutants of *Escherichia coli*. *Mol. Gen. Genet.* **214**:313–316.
  50. Efimova, E. P., E. P. Delver, and A. A. Belogurov. 1988. 2-Aminopurine and 5-bromouracil induce alleviation of type I restriction in *Escherichia coli*: mismatches function as inducing signals? *Mol. Gen. Genet.* **214**:317–320.
  51. Ellis, D. J., D. T. F. Dryden, T. Berge, J. M. Edwardson, and R. M. Henderson. 1999. Direct observation of DNA translocation and cleavage by the *EcoKI* endonuclease using atomic force microscopy. *Nat. Struct. Biol.* **6**:15–17.
  52. Endlich, B., and S. Linn. 1985. The DNA restriction endonuclease of *Escherichia coli* B. I. Studies of the DNA translocation and the ATPase activities. *J. Biol. Chem.* **260**:5720–5728.
  - 52a. Endlich, B., and S. Linn. 1985. The DNA restriction endonuclease of *Escherichia coli* B. II. Further studies of the structure of DNA intermediates and products. *J. Biol. Chem.* **260**:5729–5738.
  53. Eskin, B., and S. Linn. 1972. The deoxyribonucleic acid modification and restriction enzymes of *Escherichia coli* B. II. Purification, subunit structure, and catalytic properties of the restriction endonuclease. *J. Biol. Chem.* **247**:6183–6191.
  54. Franklin, N. C., and W. F. Dove. 1969. Genetic evidence for restriction targets in the DNA of phages lambda and phi 80. *Genet. Res.* **14**:151–157.
  55. Fuller-Pace, F. V., L. R. Bullas, H. Delius, and N. E. Murray. 1984. Genetic recombination can generate altered restriction specificity. *Proc. Natl. Acad. Sci. USA* **81**:6095–6099.
  56. Gann, A. A., A. J. Campbell, J. F. Collins, A. F. Coulson, and N. E. Murray.



1987. Reassortment of DNA recognition domains and the evolution of new specificities. *Mol. Microbiol.* **1**:13–22.
57. Garcia, L. R., and I. J. Molineux. 1999. Translocation and specific cleavage of bacteriophage T7 DNA *in vivo* by *EcoKI*. *Proc. Natl. Acad. Sci. USA* **96**: 12430–12435.
  58. Glover, S. W., and C. Colson. 1969. Genetics of host-controlled restriction and modification in *Escherichia coli*. *Genet. Res.* **13**:227–240.
  59. Gorbalenya, A. E., and E. V. Koonin. 1991. Endonuclease (R) subunits of type-I and type-III restriction-modification enzymes contain a helicase-like domain. *FEBS Lett.* **291**:277–281.
  60. Gottesman, S. 1996. Proteases and their targets in *Escherichia coli*. *Annu. Rev. Genet.* **30**:465–506.
  61. Gottesman, S. 1999. Regulation by proteolysis: developmental switches. *Curr. Opin. Microbiol.* **2**:142–147.
  62. Gough, J. A., and N. E. Murray. 1983. Sequence diversity among related genes for recognition of specific targets in DNA molecules. *J. Mol. Biol.* **166**:1–19.
  63. Gubler, M., and T. A. Bickle. 1991. Increased protein flexibility leads to promiscuous protein-DNA interactions in type IC restriction-modification systems. *EMBO J.* **10**:951–957.
  64. Gubler, M., D. Bragaglia, J. Meyer, A. Piekarowicz, and T. A. Bickle. 1992. Recombination of constant and variable modules alters DNA sequence recognition by type IC restriction-modification enzymes. *EMBO J.* **11**:233–240.
  - 64a. Hall, M. C., and S. W. Matson. 1999. Helicase motifs: the engine that powers DNA unwinding. *Mol. Microbiol.* **34**:867–877.
  65. Heitman, J. 1993. On the origins, structures and functions of restriction-modification enzymes. *Genet. Eng.* **15**:57–108.
  66. Highlander, S. K., and O. Garza. 1996. The restriction-modification system of *Pasteurella haemolytica* is a member of a new family of type I enzymes. *Gene* **178**:89–96.
  67. Hiom, K. J., and S. G. Sedgwick. 1992. Alleviation of *EcoK* DNA restriction in *Escherichia coli* and involvement of *umuDC* activity. *Mol. Gen. Genet.* **231**:265–275.
  68. Horiuchi, K., and N. D. Zinder. 1972. Cleavage of bacteriophage  $\phi$ 1 DNA by the restriction enzyme of *Escherichia coli* B. *Proc. Natl. Acad. Sci. USA* **69**:3220–3224.
  69. Hubáček, J., and S. W. Glover. 1970. Complementation analysis of temperature-sensitive host specificity mutations in *Escherichia coli*. *J. Mol. Biol.* **50**: 111–127.
  70. Hughes, S. G. 1977. Studies of plasmid-encoded restriction and modification systems. Ph.D. thesis. University of Edinburgh, Scotland.
  71. Iida, S., M. B. Streiff, T. A. Bickle, and W. Arber. 1987. Two DNA anti-restriction systems of bacteriophage P1, *darA* and *darB*: characterization of *darA*<sup>−</sup> phages. *Virology* **157**:156–166.
  72. Ives, C. L., P. D. Nathan, and J. E. Brooks. 1992. Regulation of the *Bam*HI restriction-modification system by a small intergenic open reading frame, *bamHIC*, in both *Escherichia coli* and *Bacillus subtilis*. *J. Bacteriol.* **174**: 7194–7201.
  73. Ives, C. L., A. Sohail, and J. E. Brooks. 1995. The regulatory C proteins from different restriction-modification systems can cross-complement. *J. Bacteriol.* **177**:6313–6315.
  74. Janscak, P., and T. A. Bickle. 1998. The DNA recognition subunit of the type IB restriction-modification enzyme *EcoAI* tolerates circular permutations of its polypeptide chain. *J. Mol. Biol.* **284**:937–948.
  75. Janscak, P., A. Abadjieva, and K. Firman. 1996. The type I restriction endonuclease *R.EcoR124I*: over-production and biochemical properties. *J. Mol. Biol.* **257**:977–991.
  76. Janscak, P., D. T. F. Dryden, and K. Firman. 1998. Analysis of the subunit assembly of the type IC restriction-modification enzyme *EcoR124I*. *Nucleic Acids Res.* **26**:4439–4445.
  77. Janscak, P., M. P. MacWilliams, U. Sandmeier, V. Nagaraja, and T. A. Bickle. 1999. DNA translocation blockage, a general mechanism of cleavage site selection by type I restriction enzymes. *EMBO J.* **18**:2638–2647.
  78. Janscak, P., U. Sandmeier, and T. A. Bickle. 1999. Single amino acid substitutions in the HsdR subunit of the type IB restriction enzyme *EcoAI* uncouple the DNA translocation and DNA cleavage activities of the enzyme. *Nucleic Acids Res.* **27**:2638–2643.
  - 78a. Janscak, P., and T. A. Bickle. 2000. DNA supercoiling during ATP-dependent DNA translocation by the type I restriction enzyme *EcoAI*. *J. Mol. Biol.* **295**:1089–1099.
  79. Kan, N. C., J. A. Lautenberger, M. N. Edgell, and C. A. Hutchison. 1979. The nucleotide sequence recognized by the *Escherichia coli* K12 restriction and modification enzymes. *J. Mol. Biol.* **130**:191–209.
  80. Kannan, P., G. M. Cowan, S. Daniel, A. A. Gann, and N. E. Murray. 1989. Conservation of organization in the specificity polypeptides of two families of type I restriction enzymes. *J. Mol. Biol.* **209**:335–344.
  - 80a. Kaufmann, G. 2000. Anticodon nucleases. *Trends Biochem. Sci.* **25**:70–74.
  81. Kelleher, J. E., and E. A. Raleigh. 1994. Response to UV damage by four *Escherichia coli* K-12 restriction systems. *J. Bacteriol.* **176**:5888–5896.
  82. Kelleher, J. E., A. S. Daniel, and N. E. Murray. 1991. Mutations that confer *de novo* activity upon a maintenance methyltransferase. *J. Mol. Biol.* **221**: 431–440.
  83. King, G., and N. E. Murray. 1994. Restriction enzymes in cells, not Ependorfs. *Trends Microbiol.* **2**:465–469.
  84. Klimasauskas, S., and R. J. Roberts. 1995. M.HhaI binds tightly to substrates containing mismatches at the target base. *Nucleic Acids Res.* **23**: 1388–1395.
  85. Klimasauskas, S., S. Kumar, R. J. Roberts, and X. Cheng. 1994. HhaI methyltransferase flips its target base out of the DNA helix. *Cell* **76**:357–369.
  86. Kneale, G. G. 1994. A symmetrical model for the domain structure of type I DNA methyltransferases. *J. Mol. Biol.* **243**:1–5.
  87. Kobayashi, I. 1998. Selfishness and death: raison d'être of restriction, recombination and mitochondria. *Trends Genet.* **14**:368–374.
  88. Korolev, S., J. Hsieh, G. H. Gauss, T. M. Lohman, and G. Waksman. 1997. Major domain swiveling revealed by the crystal structures of complexes of *E. coli* Rep helicase bound to single-stranded DNA and ADP. *Cell* **90**:635–647.
  89. Korona, R., B. Korona, and B. R. Levin. 1993. Sensitivity of naturally occurring coliphages to type I and type II restriction and modification. *J. Gen. Microbiol.* **139**:1283–1290.
  90. Korona, R., and B. R. Levin. 1993. Phage-mediated selection and the evolution and maintenance of restriction-modification. *Evolution* **47**:556–575.
  91. Kroger, M., and G. Hobom. 1984. The nucleotide recognized by the *Escherichia coli* A restriction and modification enzyme. *Nucleic Acids Res.* **12**: 887–899.
  92. Kruger, D. H., and T. A. Bickle. 1983. Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. *Microbiol. Rev.* **47**:345–360.
  93. Kuhnlein, U., and W. Arber. 1972. Host specificity of DNA produced by *Escherichia coli*. XV. The role of nucleotide methylation in *in vitro* B-specific modification. *J. Mol. Biol.* **63**:9–19.
  94. Kulakauskas, S., A. Lubys, and S. D. Ehrlich. 1995. DNA restriction-modification systems mediate plasmid maintenance. *J. Bacteriol.* **177**:3451–3454.
  95. Kulik, E. M., and T. A. Bickle. 1996. Regulation of the activity of the type IC *EcoR124I* restriction enzyme. *J. Mol. Biol.* **264**:891–906.
  96. Kusana, K., T. Naito, N. Handa, and I. Kobayashi. 1995. Restriction-modification systems as genomic parasites in competition for specific sequences. *Proc. Natl. Acad. Sci. USA* **92**:11095–11099.
  97. Kusiak, M., C. Price, D. Rice, and D. P. Hornby. 1992. The HsdS polypeptide of the type IC restriction enzyme *EcoR124* is a sequence-specific DNA-binding protein. *Mol. Microbiol.* **6**:3251–3256.
  98. Kutter, E., B. Guttman, and K. Carlson. 1994. The transition from host to phage metabolism after T4 infection, p. 343–346. *In* J. D. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, K. Carlson, and E. S. Miller (ed.), *Molecular biology of bacteriophage T4*. American Society for Microbiology, Washington, D.C.
  99. Labahn, J., J. Granzin, G. Schluckebier, D. P. Robinson, W. E. Jack, I. Schildkraut, and W. Saenger. 1994. Three-dimensional structure of the adenine-specific DNA methyltransferase M.Taq I in complex with the co-factor S-adenosylmethionine. *Proc. Natl. Acad. Sci. USA* **91**:10957–10961.
  100. Lautenberger, J. A., and S. Linn. 1972. The deoxyribonucleic acid modification and restriction enzymes of *Escherichia coli* B. I. Purification, subunit structure, and catalytic properties of the modification methylase. *J. Biol. Chem.* **247**:6176–6182.
  101. Lee, N. S., O. Rutebuka, T. Arakawa, T. A. Bickle, and J. Ryu. 1997. *KpnAI*, a new type I restriction-modification system in *Klebsiella pneumoniae*. *J. Mol. Biol.* **271**:342–348.
  102. Levin, B. R. 1988. Frequency-dependent selection in bacterial populations. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* **319**:459–472.
  103. Linder, P., P. F. Laska, M. Ashburner, P. Leeroy, P. J. Nielson, K. Nish, J. Schnier, and P. P. Slonimski. 1989. Birth of the D-E-A-D box. *Nature* **337**:121–122.
  104. Linn, S., and W. Arber. 1968. Host specificity of DNA produced by *Escherichia coli*. X. *In vitro* restriction of phage  $\phi$ d replicative form. *Proc. Natl. Acad. Sci. USA* **59**:1300–1306.
  105. Loenen, W. A., and N. E. Murray. 1986. Modification enhancement by the restriction alleviation protein (Ral) of bacteriophage lambda. *J. Mol. Biol.* **190**:11–22.
  106. Loenen, W. A., A. S. Daniel, H. D. Braymer, and N. E. Murray. 1987. Organization and sequence of the *hsd* genes of *Escherichia coli* K-12. *J. Mol. Biol.* **198**:159–170.
  107. Luria, S. E. 1953. Host-induced modification of viruses. *Cold Spring Harbor Symp. Quant. Biol.* **18**:237–244.
  108. Luria, S. E., and M. L. Human. 1952. A nonhereditary, host-induced variation of bacterial viruses. *J. Bacteriol.* **64**:557–569.
  109. MacWilliams, M. P., and T. A. Bickle. 1996. Generation of new DNA binding specificity by truncation of the type IC *EcoDXXI hsdS* gene. *EMBO J.* **15**:4775–4783.

110. Makovets, S., A. J. B. Titheradge, and N. E. Murray. 1998. ClpX and ClpP are essential for the efficient acquisition of genes specifying type IA and IB restriction systems. *Mol. Microbiol.* **28**:25–35.
111. Makovets, S., V. A. Doronina, and N. E. Murray. 1999. Regulation of endonuclease activity by proteolysis prevents breakage of unmodified bacterial chromosomes by type I restriction enzymes. *Proc. Natl. Acad. Sci. USA* **96**:9757–9762.
112. McKane, M., and R. Milkman. 1995. Transduction, restriction and recombination patterns in *Escherichia coli*. *Genetics* **139**:35–43.
113. Meister, J., M. MacWilliams, P. Hubner, H. Jutte, E. Skrzypek, A. Piekawicz, and T. A. Bickle. 1993. Macroevolution by transposition: drastic modification of DNA recognition by a type I restriction enzyme following Tn5 transposition. *EMBO J.* **12**:4585–4591.
114. Mernagh, D. R., and G. G. Kneale. 1996. High resolution footprinting of a type I methyltransferase reveals a large structural distortion within the DNA recognition site. *Nucleic Acids Res.* **24**:4853–4858.
115. Mernagh, D. R., L. A. Reynolds, and G. G. Kneale. 1997. DNA binding and subunit interactions in the type I methyltransferase *M.EcoR124I*. *Nucleic Acids Res.* **25**:987–991.
116. Mernagh, D. R., I. A. Taylor, and G. G. Kneale. 1998. Interaction of the type I methyltransferase *M.EcoR124I* with modified DNA substrates: sequence discrimination and base flipping. *Biochem. J.* **336**:719–725.
117. Meselson, M., and R. Yuan. 1968. DNA restriction enzyme from *E. coli*. *Nature* **217**:1110–1114.
118. Meselson, M., R. Yuan, and J. Heywood. 1972. Restriction and modification of DNA. *Annu. Rev. Biochem.* **41**:447–466.
119. Milkman, R., and M. M. Bridges. 1993. Molecular evolution of the *Escherichia coli* chromosome. IV. Sequence comparisons. *Genetics* **133**:455–468.
120. Milkman, R., E. A. Raleigh, M. McKane, D. Cryderman, P. Bilodeau, and K. McWeeny. 1999. Molecular evolution of the *Escherichia coli* chromosome. V. Recombination patterns among strains of diverse origin. *Genetics* **153**:539–554.
121. Moffat, B. A., and F. W. Studier. 1988. Entry of bacteriophage T7 DNA into the cell and escape from host restriction. *J. Bacteriol.* **170**:2095–2105.
122. Murray, N. E., J. A. Gough, B. Suri, and T. A. Bickle. 1982. Structural homologues among type I restriction-modification systems. *EMBO J.* **1**:535–539.
123. Murray, N. E., A. S. Daniel, G. M. Cowan, and P. M. Sharp. 1993. Conservation of motifs within the unusually variable polypeptide sequences of type I restriction and modification enzymes. *Mol. Microbiol.* **9**:133–143.
124. Murray, N. E., P. M. De Ritis, and L. A. Foster. 1973. DNA targets for the *Escherichia coli* K restriction system analysed genetically in recombinants between phages phi80 and lambda. *Mol. Gen. Genet.* **120**:261–281.
125. Nagaraja, V., M. Steiger, C. Nager, S. M. Hadi, and T. A. Bickle. 1985. The nucleotide sequence recognised by the *Escherichia coli* D type I restriction and modification enzyme. *Nucleic Acids Res.* **13**:389–399.
126. Nagaraja, V., J. C. Shepherd, T. Pripfl, and T. A. Bickle. 1985. Two type I restriction enzymes from *Salmonella* species: purification and DNA recognition sequences. *J. Mol. Biol.* **182**:579–587.
127. Nagaraja, V., J. C. Shepherd, and T. A. Bickle. 1985. A hybrid recognition sequence in a recombinant restriction enzyme and the evolution of DNA sequence specificity. *Nature* **316**:371–372.
128. Naito, T., K. Kusano, and I. Kobayashi. 1995. Selfish behaviour of restriction-modification systems. *Science* **267**:897–899.
129. Noyer-Weidner, M., and T. A. Trautner. 1993. Methylation of DNA in prokaryotes, p. 40–108. *In* J. P. Jost, and H. P. Saluz (ed.), *DNA methylation: molecular biology and biological significance*. Birkhauser Verlag, Basel, Switzerland.
130. O'Neill, M., A. Chen, and N. E. Murray. 1997. The restriction-modification genes of *Escherichia coli* K-12 may not be selfish: they do not resist loss and are readily replaced by alleles conferring different specificities. *Proc. Natl. Acad. Sci. USA* **94**:14596–14601.
131. O'Neill, M., D. T. F. Dryden, and N. E. Murray. 1998. Localization of a protein-DNA interface by random mutagenesis. *EMBO J.* **17**:7118–7127.
132. Patel, J., I. Taylor, C. F. Dutta, G. Kneale, and K. Firman. 1992. High-level expression of the cloned genes encoding the subunits of an intact DNA methyltransferase, *M.EcoR124*. *Gene* **112**:21–27.
133. Piekawicz, A., J. D. Goguen, and E. Skrzypek. 1985. The *EcoDXX1* restriction and modification system of *Escherichia coli* ET7: purification, subunit structure and properties of the restriction endonuclease. *Eur. J. Biochem.* **152**:387–393.
134. Powell, L. M., and N. E. Murray. 1995. S-Adenosyl methionine alters the DNA contacts of the *EcoKI* methyltransferase. *Nucleic Acids Res.* **23**:967–974.
135. Powell, L. M., D. T. F. Dryden, D. F. Willcock, R. H. Pain, and N. E. Murray. 1993. DNA recognition by the *EcoK* methyltransferase: the influence of DNA methylation and the cofactor S-adenosyl-L-methionine. *J. Mol. Biol.* **234**:60–71.
136. Powell, L. M., B. A. Connolly, and D. T. F. Dryden. 1998. The DNA binding characteristics of the trimeric *EcoKI* methyltransferase and its partially assembled dimeric form determined by fluorescence polarisation and DNA footprinting. *J. Mol. Biol.* **283**:947–961.
137. Powell, L. M., D. T. F. Dryden, and N. E. Murray. 1998. Sequence-specific DNA binding by *EcoKI*, a type IA DNA restriction enzyme. *J. Mol. Biol.* **283**:963–976.
138. Prakash-Cheng, A., and J. Ryu. 1993. Delayed expression of in vivo restriction activity following conjugal transfer of *Escherichia coli* *hsdK* (restriction-modification) genes. *J. Bacteriol.* **175**:4905–4906.
139. Prakash-Cheng, A., S. S. Chung, and J. Ryu. 1993. The expression and regulation of *hsdK* genes after conjugative transfer. *Mol. Gen. Genet.* **241**:491–496.
140. Price, C., and T. A. Bickle. 1986. A possible role for DNA restriction in bacterial evolution. *Microbiol. Sci.* **3**:296–299.
141. Price, C., T. Pripfl, and T. A. Bickle. 1987. *EcoR124* and *EcoR124/3*: the first members of a new family of type I restriction and modification systems. *Eur. J. Biochem.* **167**:111–115.
142. Price, C., J. Lingner, T. A. Bickle, K. Firman, and S. W. Glover. 1989. Basis for changes in DNA recognition by the *EcoR124* and *EcoR124/3* type I DNA restriction and modification enzymes. *J. Mol. Biol.* **205**:115–125.
143. Radding, C. M. 1973. Molecular mechanisms in genetic recombination. *Annu. Rev. Gen.* **7**:87–111.
144. Raleigh, E. A., and J. E. Brooks. 1998. Restriction-modification systems: where they are and what they do, p. 78–92. *In* F. J. de Bruijn, J. R. Lupski, and G. M. Weinstock (ed.), *Bacterial genomes: physical structure and analysis*. Chapman and Hall, New York, N.Y.
145. Raleigh, E. A., and G. Wilson. 1986. *Escherichia coli* K-12 restricts DNA containing 5-methylcytosine. *Proc. Natl. Acad. Sci. USA* **83**:9070–9074.
146. Raleigh, E. A., R. Trimarchi, and H. Revel. 1989. Genetic and physical mapping of the *mcrA* (*rglA*) and *mcrB* (*rglB*) loci of *Escherichia coli* K-12. *Genetics* **122**:279–296.
- 146a. Rao, D. N., S. Saha, and V. Krishnamurthy. 2000. The ATP-dependent restriction enzymes. *Prog. Nucleic Acids Res. Mol. Biol.* **64**:1–63.
147. Ravetch, J. V., K. Horiuchi, and N. D. Zinder. 1978. Nucleotide sequence of the recognition site for the restriction-modification enzyme of *Escherichia coli* B. *Proc. Natl. Acad. Sci. USA* **75**:2266–2270.
148. Read, T. D., A. T. Thomas, and B. M. Wilkins. 1992. Evasion of type I and type II DNA restriction systems by IncI1 plasmid ColIb-P9 during transfer by bacterial conjugation. *Mol. Microbiol.* **6**:1933–1941.
149. Redaschi, N., and T. A. Bickle. 1996. DNA restriction and modification systems, p. 773–781. *In* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed. American Society for Microbiology, Washington, D.C.
150. Roberts, R. J., and X. Cheng. 1998. Base flipping. *Annu. Rev. Biochem.* **67**:181–198.
151. Sain, B., and N. E. Murray. 1980. The *hsd* (host specificity) genes of *E. coli* K-12. *Mol. Gen. Genet.* **180**:35–46.
152. Schluckebier, G., J. Labahn, J. Granzin, and W. Saenger. 1998. M.TaqI: possible catalysis via cation-pi interactions in N-specific DNA methyltransferases. *Biol. Chem.* **379**:389–400.
153. Schouler, C., M. Gautier, S. Dusko Ehrlich, and M.-C. Chopin. 1998. Combinational variation of restriction modification specificities in *Lactococcus lactis*. *Mol. Microbiol.* **28**:169–178.
154. Scott, J. R. 1993. Conjugative transposons, p. 597–614. *In* A. L. Sonenshein, J. A. Hoch, and R. Losick (ed.), *Bacillus subtilis* and other gram-positive bacteria: biochemistry, physiology, and molecular genetics. American Society for Microbiology, Washington, D.C.
155. Sharp, P. M. 1986. Molecular evolution of bacteriophages: evidence of selection against the recognition sites of host restriction enzymes. *Mol. Biol. Evol.* **3**:75–83.
156. Sharp, P. M., J. E. Kelleher, A. S. Daniel, G. M. Cowan, and N. E. Murray. 1992. Roles of selection and recombination in the evolution of type I restriction-modification systems in enterobacteria. *Proc. Natl. Acad. Sci. USA* **89**:9836–9840.
157. Simmon, V. F., and S. Lederberg. 1972. Degradation of bacteriophage lambda deoxyribonucleic acid after restriction by *Escherichia coli* K-12. *J. Bacteriol.* **112**:161–169.
158. Sjostrom, J. E., S. Lofdahl, and L. Philipson. 1978. Biological characteristics of a type I restriction-modification system in *Staphylococcus aureus*. *J. Bacteriol.* **133**:1144–1149.
159. Smith, G. M. 1991. Conjugational recombination in *E. coli*: myths and mechanisms. *Cell* **64**:19–27.
160. Smith, H. O. 1979. Nucleotide sequence specificity of restriction endonucleases. *Science* **205**:455–462.
161. Smith, J. D., W. Arber, and U. Kuhnlein. 1972. Host specificity of DNA produced by *Escherichia coli*. XIV. The role of nucleotide methylation in *in vivo* B-specific modification. *J. Mol. Biol.* **63**:1–8.
162. Stahl, F. W., L. C. Thomason, I. Siddiqi, and M. M. Stahl. 1990. Further tests of a recombination model in which  $\chi$  removes the RecD subunit from the RecBCD enzyme of *Escherichia coli*. *Genetics* **126**:519–533.
163. Stahl, M. M., I. Kobayashi, F. W. Stahl, and S. K. Huntingdon. 1983. Activation of Chi, a recombinator, by the action of endonuclease at a distant site. *Proc. Natl. Acad. Sci. USA* **80**:2310–2313.

164. Studier, F. W., and N. R. Movva. 1976. SAMase gene of bacteriophage T3 is responsible for overcoming host restriction. *J. Virol.* **19**:136–145.
165. Studier, F. W., and P. K. Bandyopadhyay. 1988. Model for how type I restriction enzymes select cleavage sites in DNA. *Proc. Natl. Acad. Sci. USA* **85**:4677–4681.
166. Sturrock, S. S., and D. T. F. Dryden. 1997. A prediction of the amino acids and structures involved in DNA recognition by type I DNA restriction and modification enzymes. *Nucleic Acids Res.* **25**:3408–3414.
167. Suri, B., and T. A. Bickle. 1985. *EcoA*: the first member of a new family of type I restriction modification systems. Gene organization and enzymatic activities. *J. Mol. Biol.* **186**:77–85.
168. Suri, B., J. C. W. Shepherd, and T. A. Bickle. 1984. The *EcoA* restriction and modification system of *Escherichia coli* 15T<sup>-</sup>: enzyme structure and DNA recognition sequence. *EMBO J.* **3**:575–579.
169. Suri, B., V. Nagaraja, and T. A. Bickle. 1994. Bacterial DNA modification. *Curr. Top. Microbiol. Immunol.* **180**:1–9.
170. Szczelkun, M. D. 2000. How do proteins move along DNA? Lessons from type I and type III restriction endonucleases. *Essays Biochem.* **35**:131–143.
171. Szczelkun, M. D., M. S. Dillingham, P. Janscak, K. Firman, and S. E. Halford. 1996. Repercussions of DNA tracking by the type IC restriction endonuclease *EcoR124I* on linear, circular and catenated substrates. *EMBO J.* **15**:6335–6347.
172. Tao, T., and R. M. Blumenthal. 1992. Sequence and characterization of *pvuII*R, the *PvuII* endonuclease gene, and of *pvuII*C, its regulatory gene. *J. Bacteriol.* **174**:3395–3398.
173. Tao, T., J. C. Bourne, and R. M. Blumenthal. 1991. A family of regulatory genes associated with type II restriction-modification systems. *J. Bacteriol.* **173**:1367–1375.
174. Taylor, I., J. Patel, K. Firman, and G. Kneale. 1992. Purification and biochemical characterisation of the *EcoR124* type I modification methylase. *Nucleic Acids Res.* **20**:179–186.
175. Taylor, I., D. Watts, and G. Kneale. 1993. Substrate recognition and selectivity in the type IC DNA modification methylase *M.EcoR124I*. *Nucleic Acids Res.* **21**:4929–4935.
176. Taylor, I. A., M. Webb, and G. G. Kneale. 1996. Surface labelling of the type I methyltransferase *M.EcoR124I* reveals lysine residues critical for DNA binding. *J. Mol. Biol.* **258**:62–73.
177. Thoms, B., and W. Wackernagel. 1982. UV-induced alleviation of lambda restriction in *Escherichia coli* K-12: kinetics of induction and specificity of this SOS function. *Mol. Gen. Genet.* **186**:111–117.
178. Thoms, B., and W. Wackernagel. 1984. Genetic control of damage-inducible restriction alleviation in *Escherichia coli* K12: an SOS function not repressed by *lexA*. *Mol. Gen. Genet.* **197**:297–303.
179. Thorpe, P. H., D. Ternent, and N. E. Murray. 1997. The specificity of *StySKI*, a type I restriction enzyme, implies a structure with rotational symmetry. *Nucleic Acids Res.* **25**:1694–1700.
180. Titheradge, A. J. B., D. Ternent, and N. E. Murray. 1996. A third family of allelic hsd genes in *Salmonella enterica*: sequence comparisons with related proteins identify conserved regions implicated in restriction of DNA. *Mol. Microbiol.* **22**:437–447.
181. Tomb, J. F., O. White, A. R. Kerlavage, R. A. Clayton, G. G. Sutton, R. D. Fleischmann, K. A. Ketchum, H. P. Klenk, S. Gill, B. A. Dougherty, K. Nelson, J. Quackenbush, L. Zhou, E. F. Kirkness, S. Peterson, B. Loftus, D. Richardson, R. Dodson, H. G. Khalak, A. Glodek, K. McKenney, L. M. Fitzgerald, N. Lee, M. D. Adams, E. K. Hickey, D. E. Berg, J. D. Gocayne, T. Utterback, J. D. Peterson, J. M. Kelley, M. D. Cotton, J. M. Weldman, C. Fujii, C. Bowman, L. Watthey, E. Wallin, W. S. Hayes, M. Borodovsky, P. D. Karp, H. O. Smith, C. M. Fraser, and J. C. Venter. 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388**:539–547.
182. Trautner, T. A., and M. Noyer-Weidner. 1993. Restriction/modification systems in *Bacillus subtilis*, related species, and their phages, p. 539–552. In A. L. Sonenshein, J. A. Hoch, and R. Losick (ed.), *Bacillus subtilis* and other gram-positive bacteria: biochemistry, physiology, and molecular genetics. American Society for Microbiology, Washington, D.C.
183. Tyndall, C., J. Meister, and T. A. Bickle. 1994. The *Escherichia coli* *prf* region encodes a functional type IC DNA restriction system closely integrated with an anticodon nuclease gene. *J. Mol. Biol.* **237**:266–274.
184. Valinluck, B., N. Sook Lee, and J. Ryu. 1995. A new restriction-modification system, *KpnBI*, recognised in *Klebsiella pneumoniae*. *Gene* **167**:59–62.
185. Van Pel, A., and C. Colson. 1974. DNA restriction and modification systems in *Salmonella*. II. Genetic complementation between the K and B systems of *Escherichia coli* and the *Salmonella typhimurium* system SB, with the same chromosomal location. *Mol. Gen. Genet.* **135**:51–60.
186. Velankar, S. S., P. Soultanas, M. S. Dillingham, H. S. Subramanya, and D. B. Wigley. 1999. Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. *Cell* **97**:75–84.
187. Vovis, G. F., K. Horiuchi, and N. D. Zinder. 1974. Kinetics of methylation of DNA by a restriction endonuclease from *Escherichia coli* B. *Proc. Natl. Acad. Sci. USA* **71**:3810–3813.
188. Webb, J. L., G. King, D. Ternent, A. J. B. Titheradge, and N. E. Murray. 1996. Restriction by *EcoKI* is enhanced by co-operative interactions between target sequences and is dependent on DEAD box motifs. *EMBO J.* **15**:2003–2009.
189. Weiserova, M., and K. Firman. 1998. Isolation of a non-classical mutant of the DNA recognition subunit of the type I restriction endonuclease *R.EcoR124I*. *Biol. Chem.* **379**:585–589.
190. Wilkins, B. M., P. M. Chilly, A. T. Thomas, and M. J. Pocklington. 1996. Distribution of restriction enzyme recognition sequences on broad host range plasmid RP4: molecular and evolutionary implications. *J. Mol. Biol.* **258**:447–456.
191. Willcock, D. F., D. T. F. Dryden, and N. E. Murray. 1994. A mutational analysis of the two motifs common to adenine methyltransferases. *EMBO J.* **13**:3902–3908.
192. Wilson, G. G., and N. E. Murray. 1991. Restriction and modification systems. *Annu. Rev. Genet.* **25**:585–627.
193. Winter, M. 1998. Investigation of *de novo* methylation activity in mutants of the *EcoKI* methyltransferase. Ph.D. thesis. University of Edinburgh, Scotland.
194. Xu, G., J. Willert, W. Kapfer, and T. A. Trautner. 1995. *BsuCI*, a type-I restriction-modification system in *Bacillus subtilis*. *Gene* **157**:59.
195. Yuan, R., D. L. Hamilton, and J. Burckhardt. 1980. DNA translocation by the restriction enzyme from *E. coli* K. *Cell* **20**:237–244.
196. Zabeau, M., S. Friedman, M. Van Montagu, and J. Schell. 1980. The *ral* gene of phage lambda. I. Identification of a non-essential gene that modulates restriction and modification in *E. coli*. *Mol. Gen. Genet.* **179**:63–73.
197. Zinkevich, V., P. Heslop, S. W. Glover, M. Weiserova, J. Hubacek, and K. Firman. 1992. Mutation in the specificity polypeptide of the type I restriction endonuclease *R.EcoK* that affects subunit assembly. *J. Mol. Biol.* **227**:597–601.